

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:53:45 ; Search time 95.3333 Seconds
(without alignments)
1934.128 Million cell updates/sec

Title: US-09-987-701-4

Perfect score: 3649

Sequence: 1 MSISDEVNVLVRYLQESG.....LPHLVVIVPVALIELLVK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A Geneseq_23Sep04.*

- 1: Genesecp1980s.*
- 2: Genesecp1990s.*
- 3: Genesecp2000s.*
- 4: Genesecp2001s.*
- 5: Genesecp2002s.*
- 6: Genesecp2003as.*
- 7: Genesecp2003bs.*
- 8: Genesecp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3437	94.2	514	8 ADJ75513	Adj75513 Marker ge
2	3428	93.9	514	4 AAB95225	Aab95225 Human pro
3	3427	93.9	514	6 ABO07190	Ab007190 Human p53
4	3402	93.2	514	8 ADJ76263	Adj76263 Marker ge
5	3154.2	86.4	577	7 ADD14051	Add14051 Human src
6	3154.2	86.4	577	8 ADQ18019	Adq18019 Human sof
7	3069.7	84.1	459	4 ADM19860	Adm19860 Protein e
8	3064.6	84.0	542	5 ABP41760	Abp41760 Human ova
9	2753.9	75.5	700	4 ABB60376	Abb60376 Drosophil
10	2744	75.2	395	5 ABP51424	Abp51424 Human MDD
11	2643.1	72.4	584	4 ABG21351	Abg21351 Novel hum
12	2601.2	71.4	7064	7 ADG70546	Adg70546 Aspergill
13	2601.6	71.3	5635	5 ABP60991	Abp60991 Novel hum
14	2596.5	71.2	5636	7 ADJ70089	Adj70089 Human hea
15	2596.5	71.2	5636	7 ADJ83137	Adj83137 Human hem
16	2596.5	71.2	5636	8 ADK60205	Adk60205 Angiogene
17	2596.5	71.2	5636	8 ADK60506	Adk60506 Angiogene
18	2596.5	71.2	5636	8 ADP73129	Adp73129 Angiogene
19	2591.5	71.0	31267	6 ABG74786	Abg74786 Human RGS
20	2589.5	71.0	4599	8 ADI27168	Adi27168 Mouse LRP
21	2589.5	71.0	4599	8 ADI27169	Adi27169 Mouse LRP
22	2580.2	70.7	7107	4 ABB58144	Abb58144 Drosophil
23	2576.4	70.6	5509	8 ADN96832	Adn96832 Bugula br
24	2576.2	70.6	6940	8 ADN96829	Adn96829 Bugula br
25	2575.8	70.6	4599	6 ABP56837	Abp56837 Human LRP

ALIGNMENTS

RESULT 1

ADJ75513
ID ADJ75513 standard; protein; 514 AA.

XX AC ADJ75513;

DT 20-MAY-2004 (first entry)

DE Marker gene related amino acid sequence SEQ ID NO:765.

XX KW bronchial asthma; chronic obstructive pulmonary disease;

KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;

KW gene therapy; marker.

XX OS Homo sapiens.

XX PN EP1394274-A2.

XX PD 03-MAR-2004.

XX PF 04-AUG-2003; 2003BP-00254857.

XX PR 06-AUG-2002; 2002JP-00229312.

XX PA 20-MAR-2003; 2003JP-00077212.

XX PI (GENO-) GENOX RES INC.

XX DR Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;

XX WPI; 2004-193155/19.

XX PT Testing for bronchial asthma or chronic obstructive pulmonary disease by comparing the expression level of a marker gene in a biological sample from a subject with the expression level of the gene in a sample from a healthy subject.

XX PS Example 11; SEQ ID NO 765; 241pp; English.

XX CC The present invention describes a method of testing for bronchial asthma or chronic obstructive pulmonary disease. The method comprises determining the expression level of a marker gene in a biological sample from a subject, comparing the expression level determined with the expression level of the marker gene in a biological sample from a healthy subject, and judging whether the subject has bronchial asthma or chronic obstructive pulmonary disease. The marker gene comprises: (a) a group of genes (S1) whose expression levels increase when respiratory epithelial cells are stimulated with interleukin-13; or (b) a group of genes (S2) whose expression levels decrease when respiratory epithelial cells are

Aau05396 Human tit
Adq17316 Human sof
Adi21713 Rat LRP b
Aae11937 Human CG1
Abb05008 Mouse ree
Abb57065 Mouse isc
Ade60142 Rat Prote
Adn17340 V. faba e
Abb58665 Drosophil
Abb11353 Human LDL
Adn11590 Human CD9
Adn11588 Human CD9
Adn11587 Human CD9
Adn11589 Human CD9
Aar47861 Alpha 2-M
Aar60517 Human alp
Aam79091 Human pro
Aau81019 Human alp
Abp56839 Human LRP
Abu89744 Protein d

26 2575.8 70.6 26926 4 AAU05396
27 2575.8 70.6 26926 8 ADQ17316
28 2575.1 70.6 4660 8 ADI21713
29 2574.8 70.6 4636 4 AAE11937
30 2571.1 70.5 3461 5 ABB05008
31 2571.1 70.5 3461 5 ABB57065
32 2571.1 70.5 3461 7 ADE60142
33 2566.1 70.3 5825 8 ADN17340
34 2564.8 70.3 4796 4 ABB58665
35 2563.6 70.3 4563 4 ABB11353
36 2562.6 70.2 4419 8 ADN11590
37 2562.6 70.2 4419 8 ADN11588
38 2562.6 70.2 4464 8 ADN11587
39 2562.6 70.2 4464 8 ADN11589
40 2562.6 70.2 4544 2 AAR47861
41 2562.6 70.2 4544 2 AAR60517
42 2562.6 70.2 4544 4 AAM79091
43 2562.6 70.2 4544 5 AAU81019
44 2562.6 70.2 4544 6 ABP56839
45 2562.6 70.2 4544 6 ABU89744

CC stimulated with interleukin-13. Also described: (1) a reagent (I) for
CC testing for bronchial asthma or chronic obstructive pulmonary disease;
CC (2) a kit for screening for a candidate compound for a therapeutic agent
CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
CC an animal model for bronchial asthma or chronic obstructive pulmonary
CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
CC method for producing an animal model for bronchial asthma or chronic
CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial
CC asthma or chronic obstructive pulmonary disease, comprising the compound,
CC a marker gene or an antisense nucleic acid corresponding to a portion of
CC the marker gene, a ribozyme, a polynucleotide that suppresses the
CC expression of the gene through an RNAi effect or an antibody recognising
CC a protein encoded by a marker gene; and (7) a DNA chip for testing for
CC bronchial asthma or a chronic obstructive pulmonary disease, on which a
CC probe has been immobilised to assay a marker gene. (I) has respiratory
CC and antiasthmatic activities, and can be used in gene therapy. The method
CC is useful for testing for or screening for a therapeutic agent for
CC bronchial asthma or chronic obstructive pulmonary disease. The present
CC sequence is used in the exemplification of the present invention.

XX SQ Sequence 514 AA;

Query Match 94.2%; Score 3437; DB 8; Length 514;
Best Local Similarity 90.4%; Pred. No. 8.9e-52;
Matches 483; Conservative 6; Mismatches 5; Indels 40; Gaps 12;

Qy 1 MSISSDEVNVLVYRLOESGFSHSAFTFGIKSHISOSNINGALVPPAALISITIKGLQYV 60
Db 1 MSISSDEVNVLVYRLOESGFSHSAFTFGIESHSOSNINGALVPPAALISITIKGLQYV 60
Qy 61 EAEVSNEDGTLDFGRPIESLSLIDAVMPDVVQTRQAVYRDKLAQOQAAAAAASQ 120
Db 61 EAEVSNEDGTLDFGRPIESLSLIDAVMPDVVQTRQAVYRDKLAQOQAAAAAASQ 120
Qy 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFCANVPV 180
Db 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFCANVPV 180
Qy 181 SLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
Db 181 SLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
Qy 241 ATGSDGFAFIWKDGNLSTLQGHKGPFPALKWNKKNFILSAGVDKTTIWDHTGEA 300
Db 241 ATGSDGFAFIWKDGNLSTLQGHKGPFPALKWNKKNFILSAGVDKTTIWDHTGEA 300
Qy 301 KQFPFHSAALVDVWOSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360
Db 301 KQFPFHSAALVDVWOSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360
Qy 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPGTGNTNPNANMLASASF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPGTGNTNPNANMLASASF 420
Qy 421 DSTVRLVDVDRGICHTLTKHQPVSFAVSPDGRYLASGFDKCVHIMNTQ----VCLH 476
Db 421 DSTVRLVDVDRGICHTLTKHQPVSFAVSPDGRYLASGFDKCVHIMNTQ--H 478
Qy 477 -Y--LNG--QVLN-----LGR-----SICLYTLPHHLVPIPLVALIELVL-K 514
Db 479 SYRGTGIFECVNAAGDKVGASGDSVC-----V-L-D---LRK 514
RESULT 2
ID AAB95225 standard; protein; 514 AA.
XX AAB95225;
XX AAB95225;
DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:17352.
DE
XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
PA Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.
PS The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX SQ Sequence 514 AA;

Query Match 93.9%; Score 3428; DB 4; Length 514;
Best Local Similarity 90.3%; Pred. No. 1.3e-51;
Matches 482; Conservative 6; Mismatches 6; Indels 40; Gaps 12;

Qy 1 MSISSDEVNVLVYRLOESGFSHSAFTFGIKSHISOSNINGALVPPAALISITIKGLQYV 60
Db 1 MSISSDEVNVLVYRLOESGFSHSAFTFGIESHSOSNINGALVPPAALISITIKGLQYV 60
Qy 61 EAEVSNEDGTLDFGRPIESLSLIDAVMPDVVQTRQAVYRDKLAQOQAAAAAASQ 120
Db 61 EAEVSNEDGTLDFGRPIESLSLIDAVMPDVVQTRQAVYRDKLAQOQAAAAAASQ 120
Qy 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFCANVPV 180
Db 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFCANVPV 180
Qy 181 SLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
DE
XX

Db 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQVPSNKDVTSLDWNSEGTLL 240
QY 241 ATGSYDGFARIWTKDGNLSTLQGHKGPPIFALKWKKGNFILLSAGVDKTTIIWDAHTGEA 300
Db 241 ATGSYDGFARIWTKDGNLSTLQGHKGPPIFALKWKKGNFILLSAGVDKTTIIWDAHTGEA 300
QY 301 KQOFFPHSAPALDWDQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
Db 301 KQOFFPHSAPALDWDQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTNNPNANMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTNNPNANMLASAF 420
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPSPGRYLASGSPDKCVHIWNTQ----VCLH 476
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPSPGRYLASGSPDKCVHIWNTQGTALV--H 478
QY 477 -Y--LNG--QVLLN-----LGR-----SICLYTLPHHLVVIPLVALIELLVL-K 514
Db 479 SYRGTGGIPEVCWNAAGDKVGASASDGSVC-----V-L-D-----LRK 514

RESULT 3
ABO07190
ID ABO07190 standard; protein; 514 AA.
XX ABO07190;
XX DT 13-AUG-2003 (first entry)
XX Human p53 modifying protein, SEQ ID 150.
XX DE Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
XX KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
XX KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
XX KW apoptotic disorder; cell proliferation disorder.
XX OS Homo sapiens.
XX XX WO200299122-A1.
XX PN 12-DEC-2002.
XX PD 03-JUN-2002; 2002WO-US017382.
XX PF 05-JUN-2001; 2001US-0296076P.
XX PR 10-OCT-2001; 2001US-0328605P.
XX PR 15-FEB-2002; 2002US-0357253P.
XX XX (EXEL-) EXELIXIS INC.
XX XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX XX WP1: 2003-156859/15.
XX XX N-PSDB; ACD13365.
XX XX Identifying modulators of the p53 pathway for use in treating apoptotic
XX XX or cell proliferation disorders, comprises screening for agents that
XX XX modulate activity of a human ortholog of genes that modify the p53
XX XX pathway in Drosophila.
XX XX Example 2; Page 458-459; 678pp; English.

The invention relates to identifying (M1) a candidate p53 pathway
modulating agent, by contacting an assay system comprising a purified HM
polypeptide (human orthologue of genes that modify the p53 pathway in
Drosophila) or nucleic acid with a test agent under conditions, where but
for the presence of the test agent, the system provides a reference
activity, and detecting a test agent-biased activity of the assay system.
Also included are modulating (M2) a p53 pathway of a cell (comprising
contacting a cell defective in p53 function with a candidate modulator
that specifically binds to a HM polypeptide comprising an HM amino acid

sequence, where p53 function is restored), modulating (M3) a p53 pathway
in a mammalian cell (comprising contacting the cell with an agent that
specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
a disease in a patient (comprising: (a) obtaining a biological sample
from the patient; (b) contacting the sample with a probe for HM
expression; (c) comparing the results with a control; and (d) determining
whether the comparison indicates a likelihood disease). (M1) is useful
for identifying modulators of the p53 pathway. A probe for HM expression
is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
in a patient, where the cancer has greater than 25 % expression level.
Modulators identified by (M1) are useful in a variety of diagnostic and
therapeutic applications, where disease or disorder prognosis is related
to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
proliferation disorders (e.g. cancer). Another two new methods (M2 and
M3) are useful for modulating the p53 pathway of a cell, thus restoring
the p53 function of the cell, so that the cell undergoes normal
proliferation or progression through the cell cycle. (M2) and (M3) are
also useful for treating defects in the p53 pathway such as angiogenic,
apoptotic or cell proliferation disorders. The present sequence
represents a human p53 pathway modifying protein

XX Sequence 514 AA;
SQ

Query Match 93.9%; Score 3427; DB 6; Length 514;
Best Local Similarity 90.3%; Pred. No. 1.4e-51;
Matches 482; Conservative 6; Mismatches 6; Indels 40; Gaps 12;

QY 1 MTSISDEVNPLVRYLQESGFSHSAFTGKSHISOSNINGALVPPAALISIIKGLQYV 60
Db 1 MTSISDEVNPLVRYLQESGFSHSAFTGKSHISOSNINGALVPPAALISIIKGLQYV 60
QY 61 EAEVSINEDGTLPDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
Db 61 EAEVSINEDGTLPDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
QY 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPKNKAVVLRGHESEVFI 180
Db 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPKNKAVVLRGHESEVFI 180
QY 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQVPSNKDVTSLDWNSEGTLL 240
Db 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQVPSNKDVTSLDWNSEGTLL 240
QY 241 ATGSYDGFARIWTKDGNLSTLQGHKGPPIFALKWKKGNFILLSAGVDKTTIIWDAHTGEA 300
Db 241 ATGSYDGFARIWTKDGNLSTLQGHKGPPIFALKWKKGNFILLSAGVDKTTIIWDAHTGEA 300
QY 301 KQOFFPHSAPALDWDQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
Db 301 KQOFFPHSAPALDWDQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTNNPNANMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTNNPNANMLASAF 420
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPSPGRYLASGSPDKCVHIWNTQ----VCLH 476
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPSPGRYLASGSPDKCVHIWNTQGTALV--H 478
QY 477 -Y--LNG--QVLLN-----LGR-----SICLYTLPHHLVVIPLVALIELLVL-K 514
Db 479 SYRGTGGIPEVCWNAAGDKVGASASDGSVC-----V-L-D-----LRK 514

RESULT 4
ADJ76263
ID ADJ76263 standard; protein; 514 AA.
XX ADJ76263;
XX AC ADJ76263;
XX DT 20-MAY-2004 (first entry)
XX XX Marker gene related amino acid sequence SEQ ID NO:1515.

XX bronchial asthma; chronic obstructive pulmonary disease;
 KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;
 KW gene therapy; marker.

XX Mus musculus.

XX EPI394274-A2.

XX 03-MAR-2004.

XX 04-AUG-2003; 2003EP-00254857.

XX 06-AUG-2002; 2002JP-00229312.

XX 20-MAR-2003; 2003JP-00077212.

XX (GENO-) GENOX RES INC.

XX Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;

XX WPI; 2004-193155/19.

XX Testing for bronchial asthma or chronic obstructive pulmonary disease by
 PT comparing the expression level of a marker gene in a biological sample
 PT from a subject with the expression level of the gene in a sample from a
 PT healthy subject.

XX Claim 16; SEQ ID NO 1515; 241pp; English.

XX The present invention describes a method of testing for bronchial asthma
 CC or chronic obstructive pulmonary disease. The method comprises
 CC determining the expression level of a marker gene in a biological sample
 CC from a subject, comparing the expression level determined with the
 CC expression level of the marker gene in a biological sample from a healthy
 CC subject, and judging whether the subject has bronchial asthma or chronic
 CC obstructive pulmonary disease. The marker gene comprises: (a) a group of
 CC genes (S1) whose expression levels increase when respiratory epithelial
 CC cells are stimulated with interleukin-13; or (b) a group of genes (S2)
 CC whose expression levels decrease when respiratory epithelial cells are
 CC stimulated with interleukin-13. Also described: (1) a reagent (I) for
 CC testing for bronchial asthma or chronic obstructive pulmonary disease;
 CC (2) a kit for screening for a candidate compound for a therapeutic agent
 CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
 CC an animal model for bronchial asthma or chronic obstructive pulmonary
 CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
 CC method for producing an animal model for bronchial asthma or chronic
 CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial
 CC asthma or chronic obstructive pulmonary disease, comprising the compound,
 CC a marker gene or an antisense nucleic acid corresponding to a portion of
 CC the marker gene, a ribozyme, a polynucleotide that suppresses the
 CC expression of the gene through an RNAi effect or an antibody recognising
 CC a protein encoded by a marker gene; and (7) a DNA chip for testing for
 CC bronchial asthma or a chronic obstructive pulmonary disease, on which a
 CC probe has been immobilised to assay a marker gene. (I) has respiratory
 CC and antiasthmatic activities, and can be used in gene therapy. The method
 CC is useful for testing for or screening for a therapeutic agent for
 CC bronchial asthma or chronic obstructive pulmonary disease. The present
 CC sequence is used in the exemplification of the present invention.

XX Sequence 514 AA;

Query Match 93.2%; Score 3402; DB 8; Length 514;
 Best Local Similarity 89.3%; Pred. No. 4.2e-51;
 Matches 477; Conservative 8; Mismatches 9; Indels 40; Gaps 12;

Qy 1 MSISSEVNLVRYLQESGFSHSAFTGFKSHISQSNINGALVPPAALISITKGLQYV 60

Db 1 MSISSEVNLVRYLQESGFSHSAFTGFKSHISQSNINGALVPPAALISITKGLQYV 60

Qy 61 EAEVSNEDCTLFDGRIEISLIDAVPVDVQTRQAVEDKLAQQAQAAAAAASQ 120

Db 61 EAEVSNEDCTLFDGRIEISLIDAVPVDVQTRQAVEDKLAQQAQAAAAAATNQ 120

Qy 121 QGSAKXGENTANGEAGHTIANNHTDMMEVDGVEIPNKA VVLRGHESEVFCANPV 180
 Db 121 QGSAKXGENTANGEAGHTIANNHTDMMEVDGVEIPNKA VVLRGHESEVFCANPV 180
 Qy 181 SLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 Db 181 SLLVSGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 Qy 241 ATGSDGFARITWKDGNLSTLQHGKPIFALKWNKKNFILSAGVDKTTIIWDAHTGEA 300
 Db 241 ATGSDGFARITWKDGNLSTLQHGKPIFALKWNKKNFILSAGVDKTTIIWDAHTGEA 300
 Qy 301 KQOFFPHSAPALVDVQSNNTFASCSTDMCIHVCKLQDRPIKTFQCHTNEVNAIKWDPT 360
 Db 301 KQOFFPHSAPALVDVQSNNTFASCSTDMCIHVCKLQDRPIKTFQCHTNEVNAIKWDPT 360
 Qy 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKIYTIKWSPTGFTGNTNPNANMLASASF 420
 Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKIYTIKWSPTGFTGNTNPNANMLASASF 420
 Qy 421 DSTVRLWDVDRGICIHITLTKHQPVSVAFSFGDGRYLASGSPDKCVHIWNTQ----VCLH 476
 Db 421 DSTVRLWDVDRGICIHITLTKHQPVSVAFSFGDGRYLASGSPDKCVHIWNTQ--H 478
 Qy 477 -Y--LNG--QVLIN-----LGR-----SICLYTLPHHLVVIPLVALIELVL-K 514
 Db 479 SYRGTGGEIPEVCWNAAGDKVGASASDGSVC-----V-L-D---LRK 514

RESULT 5
 ADD14051

ID ADD14051 standard; protein; 577 AA.

AC ADD14051;

DT 01-JAN-2004 (first entry)

XX Human src biomarker polypeptide SEQ ID NO:240.

XX predictor set; protein tyrosine kinase activity modulator;

KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 KW gene therapy; drug sensitivity; genetic profile; cancer; human.

OS Homo sapiens.

XX WO2003062395-A2.

XX 31-JUL-2003.

XX 17-JAN-2003; 2003WO-US0001981.

XX 18-JAN-2002; 2002US-0350061P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Huang F, Fairchild CR, Lee FY, Shaw P;

XX WPI; 2003-636735/60.

XX N-PSDB; ADD14646.

XX New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein
 PT tyrosine kinase pathways.

XX Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and


```
OY 346 QGHTNEVNAIKWPTGNLLASCDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPG 405
Db 409 QGHTNEVNAIKWPSGMLLASCDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468
OY 406 TNNPNANMLASAFSTVRLWLDVDRGICHTLTIKHQPYSVAFSPDGRYLASGSPDKC 465
Db 469 TSNPNSNIMLASAFSTVRLWLDIERGVCTHTLTIKHQPYSVAFSPDGKYLASGSPDKC 528
OY 466 VHIWNTQ-----VCLH-Y--LNG--QVLLN-----LGR-----SICLYTLPHHLVIPVA 506
Db 529 VHIWNTQSNLV--HSYRGTTGGIFEVCWNRAGDKVGASADGSGVC-----V- 572
OY 507 LIELLVL-K 514
Db 573 L-D---LRK 577

RESULT 7
ADM19860
ID ADM19860 standard; protein; 459 AA.
AC
AC ADM19860;
XX
DT 20-MAY-2004 (first entry)
DE
DE Protein encoded by novel human channel/transporter gene #178.
XX
KW immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
KW cytosolic; cardiac; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW gene therapy; channel/transporter protein; rheumatoid arthritis;
KW neoplasm; cardiac arrest; cerebrovascular disorder; cerebral ischemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease;
KW ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive.
XX
OS Homo sapiens.
XX
XX WO200154472-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001307.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 26-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225256P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
```


CC germline gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences
XX
SQ Sequence 395 AA;

Query Match 75.2%; Score 2744; DB 5; Length 395;
Best Local Similarity 99.5%; Pred. No. 4.4e-39;
Matches 389; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTSSDEVNLFVRYIQESGFSHSAFTFGIKSHISOSNINGALVPPAALISIIQKGLQYV 60
DB |||||
QY 5 MSTSSDEVNLFVRYIQESGFSHSAFTFGIESHSOSNINGALVPPAALISIIQKGLQYV 64
DB |||||
QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQVTRQAYRDKLAQQQAAAAAASQ 120
DB |||||
QY 65 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQVTRQAYRDKLAQQQAAAAAASQ 124
DB |||||
QY 121 QGSAKNGENTANGENGATIANHNTDMVEVDGVEIPPNKAVILRGHSEVEFCANVPV 180
DB |||||
QY 125 QGSAKNGENTANGENGATIANHNTDMVEVDGVEIPPNKAVILRGHSEVEFCANVPV 184
DB |||||
QY 181 SLLASGSGDSTARINLSENSTSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
DB |||||
QY 185 SLLASGSGDSTARINLSENSTSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244
DB |||||
QY 241 ATGSDGCFARIWTKDGNLSTLQGHKGPFI FALKWNKKNFILSAGVDKTTIIDAHTGEA 300
DB |||||
QY 245 ATGSDGCFARIWTKDGNLSTLQGHKGPFI FALKWNKKNFILSAGVDKTTIIDAHTGEA 304
DB |||||
QY 301 KQOPFFHSAPALDVDMOSNNTFASCSTDMCIHVCKLQODRPIKTFQGHTEVNAIKWDPT 360
DB |||||
QY 305 KQOPFFHSAPALDVDMOSNNTFASCSTDMCIHVCKLQODRPIKTFQGHTEVNAIKWDPT 364
DB |||||
QY 361 GNLLASCSDDMTLKIWSMKODNCVHDLQAHN 391
DB |||||
QY 365 GNLLASCSDDMTLKIWSMKODNCVHDLQAHN 395
DB |||||

RESULT 11
ABG21351
ID ABG21351 standard; protein; 584 AA.
XX
AC ABG21351;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21342.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS85538.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 51710; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities in
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 584 AA;

Query Match 72.4%; Score 2643.1; DB 4; Length 584;
Best Local Similarity 58.1%; Pred. No. 2.4e-36;
Matches 364; Conservative 45; Mismatches 39; Indels 179; Gaps 63;

QY 2 SI-----S-----S-----D-----EV---N---FLV-YEY-LQE-SGRSHSAFTFGIKSH 33
DB |||||
QY 22 AIGAVGFRSQAASTLAGQPDGNGVCVHAGNSRIF-VPLPVGL-EVKGFSHSAFTFGIESH 79
DB |||||
QY 34 ISQSNINGALVPPAALISIIQKGLQYVEAEVSNEDGTLFDGRPIESLSLIDAVMPDVQV 93
DB |||||
QY 80 ISQSNINGTLVPPSALISILQKGLQYVEAEVSNEDGTLFDGRPIESLSLIDAVMPDVQV 139
DB |||||
QY 94 TRQAYRDKLAQQQAAA-A-A-A-AA-----AA-SQGSAXKNGENTANGENGATIA 142
DB |||||
QY 140 MRQQAFFGKLTQQAASAAATEASAKAATMTPAASIQSNPPKRNATVNGENGATIA 198
DB |||||
QY 143 NNHTDM-MEVDGVEIPPNKAVILRGHSEVEFCANVPVSDLLASGSGDSTARINLSEN 201
DB |||||
QY 199 NNHS-KPMEIDGVEIPPNKAVILRGHSEVEFCANVPVSDLLASGSGDSTARINLSEN 257
DB |||||
QY 202 STSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLLATGSDGCFARIWTKDGNLST 261
DB |||||
QY 258 SNGSGTQLVLRHCIREGGHVPNSKDVTSLDWNSDGTLAMGSDYDGFARIWTE-N-A-- 312
DB |||||
QY 262 LGQHKGPALP--KW-NKKGN--FILS-AGVDKTTIIDAHT---GEAKQOPFFHSAPAL 312
DB |||||
QY 313 -----P--ALDVQWQ--NMTFA-SCS-TDM--CI--HVCRLGCD-----H--P-- 344
DB |||||
QY 313 DVD-WQSNNTFASCSTDMCIH-VC-----KL-----G-QDRPIKTF-----QG- 347
DB |||||
QY 345 -VKTFO-GHTF--C-T--CIESICFWGGLRKLTTWTEGKLRP-KTFCSDGGLLPAGR 396
DB |||||
QY 348 --H--T-----NEVNAIKWDPITGNLLASCSDDMTLKIWSMKQDNCVHDLQ 389
DB |||||
QY 397 RPHLLTGPDIFPKRLSALFQNEVNAIKWDPSPGMLLASCSDDMTLKIWSMKQDNCVHDLQ 456
DB |||||
QY 390 HNKEIYTIKWSPTGPGTNNPNANILASASFDSTVRLWDVDRGICHTLTKHQBVPVYVA 449
DB |||||
QY 457 HSKEIYTIKWSPTGPGTNNPNANILASASFDSTVRLWDVDRGICHTLTKHQBVPVYVA 516
DB |||||
QY 450 FSPDGRYLASGSPDKCVHIWNTQ---VCLH--Y--LNG--QVLLN-----LGR-----SI 490
DB |||||
QY 517 FSPDGRYLASGSPDKCVHIWNTQSGSLV--HSYQGTGGEVFCVNCVARGDKVGSASGVS 574
DB |||||

QY	491 C-LYTLPHLHVPLVALIELLVL--K 514	Db	1492 QPLQVHDIOKESWINYSGRLNRQSLQLQVDSFLRDPSPSSVAVVAVCNIPYSMTYVERHV 1551
Db	575 CILW-----Q---LSGK 583	QY	15 -----YL-----OE-----S-----G-----FS----- 22
RESULT 12		Db	1552 IEWLDGSLTAGPDAEDWLSIROSQECSSLSALDLQOIAHOTGWQVEIWSRQFSORGG 1611
ADG70546		QY	23 -----HS-----A-----F-----T-----F-----G-----I-----K 31
ADG70546 standard; protein; 7064 AA.		Db	1612 LDALFHRHSGRDTSRALFNFTDYQGRFQSLSKWPLQRKQGEKQITLGDVSTVCK 1671
ADG70546;		QY	32 -----S-----H-----IS-----OS-----NI-----NG-----A 42
11-MAR-2004 (first entry)		Db	1672 EDLHDITWNEVVPDALEACVHDLISDTVRAQPSAICAWDGEWSYIELDLSSRLAHA 1731
Aspergillus niger nsb2 gene-encoded peptide synthetase.		QY	43 L-----VP-----P-----AA-----L-----I-----SII-----Q----- 54
Ferrichrome biosynthesis gene cluster; nsb2; peptide synthetase;		Db	1732 LAPFGVANTVVICPEKSKWTPVATLAVMKAGAASVTLDASQPLERLRSIIQTDPRVIL 1791
transgenic; ferrichrome production; siderophore; iron chelating agent;		QY	55 -----K-----GL-----QY-----V-----E-----AS-----V----- 64
black koji; malt; food product; beverage; pharmaceutical;		Db	1792 SSASKQGLGAQLTKAPNLVVDQHSISTMHITAEPLTVPDPSSKLYIVFTSGTGVKGI 1851
functional food; nutraceutical; anaemia; sideropenia; ischaemia;		QY	65 -----S-----I-----N-----ED----- 69
antianemic; vasotropic; enzyme.		Db	1852 ITHSNFSSAIRHOQKAHGFKSTRIYDFASYAFDVSWSNFIHALTVGACLICPSDEDRD 1911
Aspergillus niger.		QY	70 -----GTL-----F-----D-----G-----R----- 76
JP2003230381-A.		Db	1912 DLASLERFGATHVDMTPSAASLLPEKSPKRLTTLVVGGEKJESVESAQRWSSLSLVKNPY 1971
19-AUG-2003.		QY	77 -----P-----I-----E-----S-----L-----SLI----- 84
06-FEB-2002; 2002JP-00030145.		Db	1972 GPSECTPTATITVPTDEYKSGISGRGLGINTVNTVTDLSLVPVGVGVLLEGLPLGA 2031
06-FEB-2002; 2002JP-00030145.		QY	85 -----D-----A-----V-----M-----PD----- 90
(GEKK-) GEKKEIRAN KK.		Db	2032 GYLGDTKTAASFEVDPQLLQICQOQARHTRMYKTGDLVHVPDGSLSFVGRKDAQVK 2091
WPI; 2003-820206/77.		QY	91 -----V-----VQ-----T-----R-----Q-----Q----- 97
N-PSDB; ADG70547.		Db	2092 IHQORVELTEIESHIRTSTKIQAVALFTKSGLCANRVAVFCVIOGTGTQTAADIRLI 2151
Novel ferrichrome biosynthesis cluster comprising ornithine monooxygenase		QY	98 -----AY-----R----- 100
protein and peptide-synthetase protein, useful for producing ferrichrome		Db	2152 DPYSTLVITAYTESAKSSLDLTPAYMIPSIWIPLQHVPLSTSGKLDYKALKSWLDSMDA 2211
utilized in food or beverage products and in pharmaceuticals.		QY	101 -----D-----KL-----A-----QQ-----QA-----A-----A-----A----- 112
Claim 2; SEQ ID NO 3; 60pp; Japanese.		Db	2212 KTFANILITASDGDVLRKAETELEQVIVEACAKILNITASKVNLDRSFANGGDSISAMR 2271
The invention relates to the Aspergillus niger ferrichrome biosynthesis		QY	113 -----A-----A-----AA-----AS----- 119
gene cluster which comprises the nsb1 gene (ADG70545) encoding an		Db	2272 LVAHCRADVNVFSAKLKSKTLAALASSKIKSASNVILGFVEEKSDFSALSPIOQWFFE 2331
ornithine monooxygenase (L-ornithine N5-oxygenase; ADG70544) and the nsb2		QY	120 Q-----Q-----QG-----SA-----K-----NG-----E-----NT----- 130
gene (ADG70547) encoding a peptide synthetase (ADG70546). The invention		Db	2332 QGLYKRSNDNFQNGFYLVKKEPLITKDIDSALSKVQVHHMLRFRHNGDEWTKTLK 2391
also relates to recombinant vectors, host cells and transgenic organisms		QY	131 -----ANG-----E-----E-----N-----G-----AH----- 139
comprising at least one of the ferrichrome biosynthesis gene cluster		Db	2392 PDTNGLYHPGVHHTCLPADIERLALSRHQMDIEKGPVFSADICHNAFGEQVLLIAHL 2451
coding regions; and method for protein engineering of a peptide		QY	140 ----- 144
synthetase involving domain rearrangement. Transgenic organisms		Db	2452 VVDLSVRVILEDBSLGGSNLQPSLPFQVWMDMQIERAKESLLDPENVLSTGTGINN 2511
comprising all or part of the ferrichrome biosynthesis gene cluster		QY	145 -----H-----T-----D-----M-----ME-----VD----- 152
(e.g., Escherichia coli or Aspergillus sp.) are able to produce large		Db	2512 LDFWQATATKMTVEDHLNFCFKIDSSKSSKQITRSTLEPVDLLLLAAVWHAFFKTF 2571
amounts of ferrichrome using black koji (malt) as a substrate.		QY	153 -----G-----DVE-----I-----P----- 158
Ferrichrome is a safe iron-chelating agent (siderophore) which is widely		Db	2572 PQDGLTIFIEGHGEPWSSDIDLSTVCGWFTTISPIHVSKSDVHKSVASLVVRVVKDARR 2631
used in pharmaceuticals or in food or beverage products. It can be used			
for the amelioration of anaemia (particularly in a functional food) and			
for treating sideropenia and ischaemia. The present sequence represents			
the Aspergillus niger ferrichrome biosynthesis gene cluster nsb2 gene-			
encoded peptide synthetase.			
Sequence 7064 AA;			
Query Match	71.4%; Score 2605.2; DB 7; Length 7064;		
Best Local Similarity	9.1%; Pred. No. 2.2e-30;		
Matches	453; Conservative 48; Mismatches 6; Indels 4448; Gaps 368;		
QY	1 M-----S-----S-----I-----S-----S-----S----- 5		
Db	1432 MEQVELELLVDPAFTSLVDPRGLVEHVLEILLPKRLKSTNELSAYRYAAVHLKSNQLA 1491		
QY	6 -----D-----E-----VN-----FL-----VY-----R----- 14		

Db	4787	LAENKIELASSEHDQLINEHIEALQSLRLRESLPQYMIPLSVWLVLNLPMTASGQDNKAL	4846
Qy	383	-----C-----	383
Db	4847	KSWLENWDETLP SKINNANGSDIIRKPDTEDE RVLSQKCISVILNMPVDKINLKD\$FIANG	4906
Qy	384	-----V--H-D---L---	387
Db	4907	GDSISAMRLASHYRTVGISIVSTLLQS KTLADPAAPSGATAISGVSQEHTDVPFELSP	4966
Qy	388	-----OO-H---NKE-IX---T-I---	397
Db	4967	IQQWFDPQSPMQCKDHREYN-QGFYVRURRTVRINDLES AFLSLVNRHAMLRSRFQH	5025
Qy	398	--KW-----S-----P-----	401
Db	5026	GGRWKOIILSHXRALHLNVSOHLMS EIASLAQERHQIDIEKPVPFSVDICLLGOQQH	5085
Qy	402	-----T---G-----P-----	404
Db	5086	LVMIAHHLVTDLVSWRIILDLE TLINGHS LTAALPFQVWSRLQAERAVS TLKPHNLLS	5145
Qy	405	--GT-NV-----PN-----AN-----LML----	415
Db	5146	TDCGVHNKLFPWKYTHDTNPCLADHRLRSVTIDRET TAVLLGEANNAMNTPVEVILL SAVW	5205
Qy	416	-A-----SA-S-P-----D-ST-----VRL	426
Db	5206	DAPFRFTSQRNSUTIFNEGHGREAWS DEIDLUS TVGWFTTLSPINIYRNATSETDM VRL	5265
Qy	427	-----W-----DVD	430
Db	5266	VKDARSLPANGWSYFTSRVLPDGPAPESHNTVSEVVNYHGQFOQLESHQALPED ID	5325
Qy	431	-----R---G-----IC-	434
Db	5326	LUGRVQRGRSISAGSUFNIEVAIEAMQAHFEFSVNQNIAHQSLINOWIDIQIPSLERICL	5385
Qy	435	I-----HTL-----T-----	439
Db	5386	VLEANPTHLCDFKFI SLDYQRLDDLTSRLLP EIESINOSTVEEIPSCSPIVDGMLLSQ	5445
Qy	440	-----K-----H-----Q---EP-V-----Y---	446
Db	5446	IKQPESYKTQRYEVLLSHDHPICLDTLKIAWQRVISRQPALRTVFIAGLDGSTAFYQAL	5505
Qy	447	-----S---V-----AFS-----PD-----	453
Db	5506	LKQCSGVI VVEAKTEEBEALKAFSSLPKV DYOQAKPPHRTLCTQTPDDKVFVCOIEMSHA I	5565
Qy	454	-----G-----G-----	454
Db	5566	TDGASSTILIKOLIDAYGDRLSSTD LVKTTREFASHLLAKPQSKISYMNTKLKGLEPCR	5625
Qy	455	-----R-----	455
Db	5626	FPSLSMSRKHCSESIEGVFVEDKMPAQIODFC SINQVTPASLLKSAWALTISTYVQNG	5685
Qy	456	-----YLASG-----	460
Db	5686	SVCFGYLASGRDLPIAGMDES VGAYTNIMVCRADLDGQPGVALVRQLONQLMODLSFQH	5745
Qy	461	-----SF-----D-----K-----	464
Db	5746	ISLASIOHELG LASDOOLFNSIVSFORS GDDNEQSAPEGKLRPKNIDGLDPTDYDIVLGI	5805
Qy	465	-----C-----V-H-----I-WN-T	471
Db	5806	NQOTRSIEDLFPSHCSLTSNQAKRILEHLOSNIAAILHNEPPALISPQDEQDIWNST	5865
Qy	472	QV-----CLH-----Y-----L-N-----	479

[illegible]

PR 30-JAN-2001; 2001US-0264922P.
 PR 06-FEB-2001; 2001US-0266797P.
 PR 19-MAR-2001; 2001US-0276988P.
 PR 04-APR-2001; 2001US-0281335P.
 PR 08-MAY-2001; 2001US-0289622P.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Agarwal P, Birkeland M, Cogswell JP, Kabnick KF, Lai Y;
 PI Martensen SA, Rizvi SK, Smith RF, Strum JC, Xie Q;
 XX WPI; 2002-508784/54.
 DR N-PSDB; ABQ86156.
 XX
 PT Secreted proteins and polynucleotides useful as vaccines for preventing
 PT or treating various diseases e.g. cancer, wounds, atherosclerosis,
 PT Parkinson's disease, Alzheimer's disease, infection, autoimmune disorder.
 XX
 PS Claim 1(a); Page 285-297; 335pp; English.
 XX
 CC The invention relates to an isolated polypeptide with signal sequences
 CC which allow it to be secreted extracellularly or membrane associated. The
 CC activity of polypeptides of the invention may be described as,
 CC cytostatic, vulnerary, antiarteriosclerotic, antiparkinsonian, neurotropic,
 CC neuroprotective, immunosuppressive, haemostatic, antiinflammatory, anorectic,
 CC cardiant, antitumor, virucide, antithyroid, cerebroprotective, anorectic,
 CC and metabolic. Polypeptides and polynucleotides of the invention are
 CC useful in the treatment, or as a vaccine in the prevention of, cancer,
 CC wound healing disorders, infection, atherosclerosis, Parkinson's disease
 CC and Alzheimer's disease, autoimmune disorder, haematopoietic disorder,
 CC inflammation, neoplastic diseases, nervous system related disorders and
 CC cardiovascular disorders, pancreatitis, respiratory disorder,
 CC hyperproliferation, systemic autoimmune disease, hyper-immunity,
 CC developmental abnormality, gastrointestinal ulceration, neuropathy,
 CC haematological diseases, metabolic diseases, sperm dysfunction, thyroid
 CC disorders e.g. hypothyroidism, brain damages, colitis, cone photo-
 CC transduction deficiency, neurological diseases, stroke, angiogenesis,
 CC ovulation disorders, diseases in the spinal cord, thyroid gland, heart,
 CC trachea, thymus, lymph node and muscular system, obesity, anorexia,
 CC growth abnormalities, and alleviation of precocious puberty. The
 CC sequences given in records ABP60965-ABP61019 represent novel human
 CC proteins of the invention
 XX
 SQ Sequence 5635 AA;

Query Match 71.3%; Score 2601.6; DB 5; Length 5635;
 Best Local Similarity 9.8%; Pred. No. 8.5e-31;
 Matches 431; Conservative 67; Mismatches 13; Indels 3904; Gaps 372;
 QY 1 M-S-----I-----SS--D-----EV----- 8
 Db 1 MISWEVVHTVFLFALLYSLAQAQSPQSEIRABEIPEGASTLAFVFDVTGSMYDDLQVVI 60
 QY 9 -----NF-LV----- 12
 Db 61 EGASKILETSLKRPKPLNFALVPFHDPEIGPVTITDPKKFQYELRELYVQGGDCPE 120
 QY 13 -----Y-----R-----Y-----LQ----- 17
 Db 121 MSIGAIAKIALEISLPGSFYVFTDARS KDYRLTHEVLQIQKQSQWFLTGDCDDRTH 180
 QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
 Db 181 IGKVVETIASTSGQVHLDKKQVNEVLKWEAEVQASKVHLLSTDHLEQAVNTWRIFP 240
 QY 27 -----T-----FG-----I-----K 31
 Db 241 DPSLKEVTVSLGSPSMIEIRNPLGLIKKGFGLHELLHNHSAKVNVKPEAGWTVK 300
 QY 32 ---S--H-----I-----S-----Q----- 36

Db 301 TSSSGRHSVRIITGLSTIDFRAGFSRKPTLDFKKTVSRRPVQGIPTVYLLNTSGISTPARID 360
 QY 37 -----S-----NI-----N-----G----- 41
 Db 361 LLELLSISGSSSLKTIPIVKYPPHRKPYGIWNISDFVPPNEAFFLKVTDYDKDKDLFORVSS 420
 QY 42 -----A-----L-----VP-----P-----A 47
 Db 421 VSFSSIVDPAPKVTMPKTPGYLQPGQIPCSVDSLLPFTLSFVRNGVTGLVDQYKLESA 480
 QY 48 -----A-----L-----I-----S-----IIQ-----KG----- 56
 Db 481 SVNLDIAKVTLSDEGFEICIAVSSAGTGRAQTFFDVSEPPPIQVPPNVTVTGERAVLT 540
 QY 57 -----L-----Q-----Y-----V-----E----- 61
 Db 541 CLIIASVDYNLTWQRNDRVRLAEPARTLANLSLELKVRFNDAGEYHCVMSSEGGSS 600
 QY 62 -A-----EVS-----NE----- 68
 Db 601 AASVFLTVQEPKVTVMKQSQFTGSGSEVSMCSATGYPKPKIAWTVNDMFIVGSHRYRM 660
 QY 69 --DGLTF-----D-----G-----R-----P-----I-- 78
 Db 661 TSDGTLFIKNAAPKDAIGYGLASNSAGTDKQNSTLRYIEAPKLMVVSSELLVALGDITV 720
 QY 79 -E--S-----L--S--LI-----DA-----V----- 87
 Db 721 MECKTSGIPPPQVKNFKGDLRLPSTFLIIDPLGLLKIQTQDLDAGDYTCVAINAAGR 780
 QY 88 -----M-----P-DV-----VQ-----T----- 94
 Db 781 ATGKITLDVGSPPVFTQEPADVSMEIGSNVTLPCYOGYPEPTIKWRRLDNMPISRPSPS 840
 QY 95 -----R----- 95
 Db 841 VSSISQRTGALFILNLWASDKGTVICEAENQFKIQSETTVTGTGLVAPLIGISPSVAN 900
 QY 96 ---QQ-----A-----Y-----R-----D--KL-- 103
 Db 901 VIEGQULTPCTLLAGNPPIPERWIKNSAMLLQNPYITVRSGLSHIERVQLQDGEYTC 960
 QY 104 -----A-----Q--QQ-----AA----- 109
 Db 961 VASNVAGTNKTTSVVHVLPITIQHQQLLSTIEGIPVTLPCKASGNPKPSVWSKKGEL 1020
 QY 110 ---AAA---A--A-----A--AA--A----- 118
 Db 1021 ISTSSAKFSAGADGSLYVSPGGEESGEYVCTATNTAGYAKRKVQLTVYVRPVFGDQRG 1080
 QY 119 -SQ-----Q-----Q-----GS----- 123
 Db 1081 LSQDKPVEISVLAGEVTLPCVKSLPPIITWAKETQLISPFSPRHTFLPSCSMKITET 1140
 QY 124 -----A-----A-----K-----N----- 126
 Db 1141 RTSDSGMYLVATNAGNVQAVKLVNHPVKIQRGPKHLKVQGVQGVQVDPICNAQGTPLP 1200
 QY 127 -----GE--N-----T--A--N--G----- 133
 Db 1201 VITWSKGGSTMLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260
 QY 134 -----E--N-----E--N-----CA-----H----- 139
 Db 1261 PPTVEDLEPPYNTTQERVANQRIEPPCPAKTPKPTIKWLNHRELGTREPGISILEDG 1320
 QY 140 T--IA-----N-----N-----HT-----DM----- 148
 Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHPPVPIKDKQVTVNSVLLNQLTNL 1380
 QY 149 -----M-----E----- 150
 Db 1381 FCEVEGTFPIIMWYKDNVQVTTESSTIQTVNNKILKLFRTATPEDAGRYSCAKINIAIGTS 1440

Qy	151	----	VD	-----	G	-----	DV	-E-	-----	I	-----	P	-----	PN	160
Db	1441	QKYNIDVLVPPTII	IGTNFPNEVS	VWLNRDVALE	CQVKGTFFPD	IHWFKDGLFLGDPN	1500								
Qy	161	----		-----	K	-A-	-----	V	-----	V	-----	VL	165		
Db	1501	VELDRGOVLHKNARND	KGRYQCTVSN	AAGKAQAKDIKI	TINPPSIKGNVT	DISVL	1560								
Qy	166	----	RG	-----									H	168	
Db	1561	INSLIKLECETRGLPMP	AITWYKDGFIMSSQA	LYDKGYLHI	PRAQVSDSATYCHV	1620									
Qy	169	----	E-S	-----	E	-----	VPI	-C-A-	-----	W	-----	177			
Db	1621	ANVAGTAESFHDVVYP	PMIEGNLATPLNKQ	VVIAHSHTLECA	AGNPSPILTWKDG	1680									
Qy	178	----	N	-----									P-V	SD-182	
Db	1681	PVRANDNIRIEAGGKKU	LEIMSQBIDRGVIC	ATSVAGEKEIKYE	VDVLPVPAIGDE	1740									
Qy	183	----	LL	-----	ASGS	-----	G	-D-	-----	S	-----	S	191		
Db	1741	TSYFIWMNNLLEDCH	TVTSGPPTIMWLKDG	QLIDERDGFKILLNG	RKLIVAAQVSNT	1800									
Qy	192	----	TA	-----									I-W	196	
Db	1801	GLYRCMAANTAGDHKE	FVTHVHPPTIKSSGL	SERVVVKYPVALQC	IANGIPNSITW	1860									
Qy	197	----	NL	-----									S	199	
Db	1861	LKDDQPNTAQNLKIO	SSGRVLQIAKTLED	AGRYTCVATNAAG	EQHQIHLHVHEPS	1920									
Qy	200	-E-	N	-S-											
Db	1921	LEDAGMLNETVLVSN	PVQLECKAAGNP	VPVIWTYKDNRLLS	GSTMTFLNRQIIDIES	1980									
Qy	208	-Q-	L	-V-											
Db	1981	AQISDAGIYKCVAINS	AGATELFYSYLQHVAP	PSISGSNNMVAVVN	PVRECEARGIPA	2040									
Qy	216	----		-----	R	-----	E	-G-	-----	G-Q-D-	-----	221			
Db	2041	PSUTWLKDGSPVSFS	NGLQVLSGGRILAT	SAQISDTGRYTCA	VAVNAAGEKORDLRV	2100									
Qy	222	-VPSN	-----												
Db	2101	YVPNIMGESQNVS	LISOAVELLCOSDAI	PPPRTLWLKDGHP	LLKPKGLSISENSVLK	2160									
Qy	232	----	D	-----											
Db	2161	IEDAQVDTGRYTCE	ATNVAGTEKNVNVNI	WPNNIGSGDELQT	LTVEGNLISLCS	2220									
Qy	242	T	----	GS-Y-D	-----	G	-----	F	-----	A	-----	249			
Db	2221	SGIPPPNLIWKKKG	SPVLTDMSGRVRI	LSGGRLQISIAEK	SDAALYSVASNVACTAKK	2280									
Qy	250	----		-----	R	-----	I	-----	WTQG	-----	256				
Db	2281	EYNLQVIRPTINSG	HPTEIIIVTRGKS	ISLECEVOGIPP	PTVTWMOGHPLIKAGVE	2340									
Qy	257	----	NL	-AS-											
Db	2341	ILDEGHILQKNIH	SVDTGRYCVAVNV	AGMTDKKYDLSH	APPSIIGNHRSPENISVVE	2400									
Qy	266	K	----	G-P-I	-FAIK-W-										
Db	2401	KNSVSLTCEASGI	PLPSITWF--	KDGWPVLSNS	VRILSGRRMLRMQ	TMEDAGQYTCV	2458								
Qy	275	----		-----	N	-----	K-K	-----	GN	-----	279				
Db	2459	VRNAAGEERKIFGL	SVLPPPHIVGENT	LEDVKVKEQS	VTLTCEVTGNPVPEITWHKQG	2518									

QY	280	-----F-I-----L-----S-----AGV-----286
Db	2519	PLQBEAHHISGGRFLQITNVQVPHTGRYTCCLASSPAGHKSRFSFLNVEVSPTIAGVGS 2578
QY	287	D---K-T-----T-I-----I-W-----293
Db	2579	DGNFEDVTILNSFTSLVCRAYSYPPATITWFKDGTPLESNRIRILPGORTLQILNAAQE 2638
QY	294	D-----A-----H-----TG-----EA 300
Db	2639	DNAGRYSCVATNEAGEMIKHVEVKVIYPIINKGDLWGPGLSPEVKIKVNNLTLTLECEA 2698
QY	301	-----K-----Q-----Q-----F 304
Db	2699	YAIPSASLSMYKDGQPLKSDDHVIAANGHTLQIKEAQISDGTGRYTCVASNIAGEDELDF 2758
QY	305	-----P-F-----H-----H-----H-----307
Db	2759	DVNIQVPSFOKLWEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPPTLTWYKQGHPL 2818
QY	308	SA-----P-----A-----LDV-----D-315
Db	2819	TSSDKVILPGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPILKIGANSDL 2878
QY	316	-----WQ-----SN-----319
Db	2879	PEEVTVLNKSALTICLSSGSPAPRNSWQDGQPLEDDHHKLSNGRILQILNTQITDI 2938
QY	320	-----NT-----PAS-C-----S-326
Db	2939	GRYVCVAENTAGSAKKYFNLNHVHPPSVIGPKSENITVVVNNFISLTCEVSGFPPLSLW 2998
QY	327	T-----D-----MC-----331
Db	2999	LKNEQIKLNTNTLIVPGRFTLQIIRAKVSDBGYTCIAINQAGSEKKFSLTVYVPPSI 3058
QY	332	--H-----V-C-----K-----L-GQ-----338
Db	3059	KDHSESLSVNVREGTSVSLCESNAVPPVITWYKNGRMITESTHVEILADGQMLHIK 3118
QY	339	-----D-----R-----P-IK-----TF-----345
Db	3119	KAESDGTQYVCRAINVAGRDDKNFHLNVYVPSIEGPEREVIVETISNPVTLTCDATGI 3178
QY	346	-----O-----G-----H-----T-----349
Db	3179	PPPTIAMLKNHKRIENSLSLEVRILSGSKLOIARSOHSDSGNYTCIASNMECKAQYF 3238
QY	350	-----N-E-V-NA-----IKW-D-P-----T-----360
Db	3239	LSIQPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIQWLKDGKPIASGETERIRVS 3298
QY	361	-G--NLL-A-S-----C-S-----D-----D-----MTL-----373
Db	3299	ANGSTLIYGALTSDTGKYTCVATNPAGEEDRIFNLNVYVTPTRIGNKDEAEKMLTVD 3358
QY	374	-K-----I-W-----SM-----KQ-D-D-----NCV-----384
Db	3359	SINTECRATGTPPPQINWLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVVTCVASNR 3418
QY	385	-----H-----D-----D-----L-Q-----388
Db	3419	AGVDNKHNLQVAPPNMDSMGTEBITVLKSGTSMACITDGTTPAPSMALWRDGPGLG 3478
QY	389	-----Q-----H-N-----H-N-----KEI 394
Db	3479	DAHLTVSTHGMVLQQLKAETEDSGKYTCIASNEAGVSKHFILKVLPEPHINGSERHEE 3538
QY	395	-----YT-----I-K-W-----KQ-D-D-----399
Db	3539	SVIVNNPLETCIASGIPAPKMTWMKDGRLPQTDQVLTGGGEVLRIISTAQVEDTGRYT 3598
QY	400	-----SPTG-----P-GT-----N-----N-----N 408

```

Db 3599 CLASSPAGDDKEYLVVRVHPNPNTAGTDEPRDITVLNRQVLTLECKSDAVPPPVITWLRN 3658
Qy 409 -----P-----N-----ANLM-LAS--A---S--F----- 420
Db 3659 GERLQATPRVRLSGRYLQINNDLGDGTANTYTCVAGNIAGKTTREFILATVNVPPNIKGG 3718
Qy 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433
Db 3719 POSLIVILLNKSTVLECTAEGVPTTRITWRKDGAVLAGNHARYSILENGFLHIQSAHVTD 3778
Qy 434 -----C-----I-----H-----TL----- 439
Db 3779 GRYLCMATNAAGTDRRIDQLQVHPPSIAPGPTNMTVIVNVQTTLACEATGIPKPSINWR 3838
Qy 440 K--H-----Q--E-----P--V-----Y-----S 447
Db 3839 KNGHLNVNDQNSRYLLSSGSLVIIISFSDDTATYECTVTNGAGDDKRTVDTLVQVPPS 3898
Qy 448 VA-----F--S-----P--D----- 453
Db 3899 IADEPTDLTKHAPAVITCTASGVPFSPSIHWTNKGIRLLPRGDGYRILSSGAIEILATQ 3958
Qy 454 -----GRY-----L---ASG--S--F 462
Db 3959 LNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIOQPSELHVILNPNILLPCEATGTPSPF 4018
Qy 463 -----L--G-----D-----KCV----- 466
Db 4019 ITWQKGINVNTSGRNHVLPSGGLQISRAVRDAGTYMCVQONPAGTALGKIKLVQVP 4078
Qy 467 -----H-----I--W-----NT--Q--V 473
Db 4079 PVISPHLKEVIAVDKEITLSCADGLPPDITWHKDGRAIVESIRORVLSGSLQIAFV 4138
Qy 474 -----C-----L-----HY----- 477
Db 4139 QPGDAGHYTCMAANVAGSSSTSKLTIVHPPIRSTEGHYTVNENSOAILPCVADGIPTP 4198
Qy 478 -----L--N-----GO-----VL-----L-----N----- 485
Db 4199 AINWKKDVLNALLGKYTAEPYGELENVVLEDSGFYTCVANNAAGEDTHTVSLTVHV 4258
Qy 486 -----L--G-----R--SIC-----L--VT-----LP--H-----H--LV--- 500
Db 4259 LPTFTLPGDVLNKGEOILRLS-CKATGIPLPKLTWTTFNNIIPAHFDSVNGHSELVIER 4317
Qy 501 -----V--I-----P-----L-----V-----A--L----- 507
Db 4318 VSKEDSGTYVCTAENS VGFKAGFVYVKEPPVFKGDYPSNWIEPLGGNAILNCEVKGDP 4377
Qy 508 -----IE-----L-----L--VL-----K 514
Db 4378 TPTIQNRKGVDEISHRIQLNGSLAIYGTVNE 4412

```

RESULT 14

ADJ70089

ID ADJ70089 standard; protein; 5636 AA.

XX AC

XX AC

XX ADJ70089;

DT 06-MAY-2004 (first entry)

DE XX

XX Human heat mitochondrial protein as a therapeutic target SeqID1895.

KW mitochondrial; human; screening assay; diabetes mellitus;

KW Huntington's disease; osteoarthritis;

KW Leber's hereditary optic neuropathy; LHON;

KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;

KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;

KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;

KW osteopathic; ophthalmological; cytostatic.

XX OS Homo sapiens.

XX PN WO2003087768-A2.

XX PD 23-OCT-2003.

XX PF 04-APR-2003; 2003WO-US010870.

XX PR 12-APR-2002; 2002US-0372843P.

XX PR 17-JUN-2002; 2002US-0389987P.

XX PR 20-SEP-2002; 2002US-0412418P.

XX PA (MITO-) MITOKOR.

XX PA (BUCK-) BUCK INST AGE RES.

XX PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

XX PI Warnock DE;

XX PI WPI; 2003-845369/78.

XX PT Identifying a mitochondrial target for drug screening assays and for

XX PT treating diseases associated with altered mitochondrial function,

XX PT comprises detecting a modified polypeptide in a sample and correlating

XX PT with the disease.

XX PS Claim 1; SEQ ID NO 1895; 180pp; English.

XX CC This invention relates to novel mitochondrial targets that can be used

XX CC for therapeutic intervention in treating a disease associated with

XX CC altered mitochondrial function. Specifically, it refers to a method for

XX CC identifying proteins of the human heart mitochondrial proteome that are

XX CC useful for drug screening assays, as well as therapeutic targets. The

XX CC present invention describes a method for identifying such proteins that

XX CC can be used in the treatment of various diseases associated with altered

XX CC mitochondrial function including diabetes mellitus, Huntington's disease,

XX CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial

XX CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy

XX CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these

XX CC compositions have neuroprotective, nontropic, antidiabetic,

XX CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and

XX CC cytostatic activities. This polypeptide sequence is a human heart

XX CC mitochondrial protein of the invention.

XX SQ Sequence 5636 AA;

Query Match 71.2%; Score 2596.5; DB 7; Length 5636;

Best Local Similarity 9.7%; Pred. No. 1.1e-30;

Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

Qy 1 M-S-----I-----SS--D-----EV----- 8

Db 1 MISWEVHTVFLFALLYSLAQDASQSEIRAEPEEGASTLAFVFDVTGSMYDDLQVVI 60

Qy 9 -----NF-LV----- 12

Db 61 EGASKILETSLRKPKRPLENFALVPFHPDPEIGPVITTDPKKPYRELRELYVQGGDCPE 120

Qy 13 -----Y-----R-----Y-----LQ----- 17

Db 121 MSGAIKIALEISLPGSFYVFTDARKSDYRLTHEVLQIQKQSQVWFVLGTDCDDRTH 180

Qy 18 -----E-----SG--F-----S--H--S-----A-----F 26

Db 181 IGKVEEIASTSSGQVFLDKKQVNEVLKVVBEAVQASKVHLLSTDHLEQAVNTWRIFP 240

Qy 27 -----T-----PG-----I-----K 31

Db 241 DPSLKEVTVSLSGPSPMIEIRNPLGKIKGGLHLLHNHSKVVNVKEPEAGNWTVK 300

Qy 32 -----S--H-----I-----S--Q----- 36

Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKRPTLDFKKTVSVPVQGIPTVYLLNTSGISTPARID 360

QY 37 -----S-----NI-----N-----G----- 41
Db 361 LLELLSISGSSLKTIPVKYYPHRKPYGIWNISDFVPNEAFFLKVTGYDKDDYLFQVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
Db 421 VSFSSIVDPAPKVTMEKTPGYVLOPQIPCSVDLSLPTLSFVRNGVTGLQVQYLKESA 480
QY 48 -----A-----L-----I-----S-----IIQ-----KG----- 56
Db 481 SVSLDIKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPVVIQVNNVTVTGERAVLT 540
QY 57 -----L-----Q-----G-----R-----P-----Y-----V-----E----- 61
Db 541 CLIIISAVDYNLTWQRNDRVRLAEPARIRTLANLSLELKSVEKFNDAEYHCHVSSSGSS 600
QY 62 -A-----EVSII-----NE----- 68
Db 601 AASVFLTVQEPKVTYMPKNQSFSGSEVMSATGYPKPKIAMTVNDMFIVSHRYRM 660
QY 69 -D-----D-----G-----R-----P-----I----- 78
Db 661 TSDGTLFIKNAAPKADAGIYGLASNSAGTDKQNSTURYEAPKLMVQSELLVALGDIIV 720
QY 79 -E-----S-----L-----S-----LI-----DA-----V----- 87
Db 721 MECKTSGIPPPQVWFKGDLRLPSTFLIDPLGLLKIOETQDLADGYTCVAINAEGR 780
QY 88 -----M-----P-----DV-----VO-----T----- 94
Db 781 ATKITLGVSPVFIQEPADVSMIEGNSVTLPCYVQGYPEPTIKWRDLNMPIFSRPFS 840
QY 95 -----R----- 95
Db 841 VSSISQRTGAULFILNLWASDKGTIYCEAENOPGKIQSETTVTVTGLVAPLIGISPSVAN 900
QY 96 -----QO-----A-----Y-----R-----D-----KL----- 103
Db 901 VIEGQQLTLPCTLLAGNPPIERWIKNSAMLLQNPIYTVRSDSGLHIERVOLQDGGEYTC 960
QY 104 -----A-----Q-----QO-----AA----- 109
Db 961 VASNAVGTNNKTVSVVHVHPTLTIHQQLSTIEGIPVTLPCKASGNPKPSVIWSKKGEL 1020
QY 110 ---AAA---A-A-----A-AA-A----- 118
Db 1021 ISTSAKFAGADGSLYVVSPEGESYVCTATNTAGYAKRKVQLTVYVRPRVFGDLRG 1080
QY 119 -SQ-----GE-----N-----Q-----GS----- 123
Db 1081 LSQDKPVEISVLAGEBVTLPCEVKSLLPPIITWAKETQLISPPSPRHTFLPSGSMKITET 1140
QY 124 -----A-----A-----K-----N----- 126
Db 1141 RTSDSGMVLCVATNIAGNVTAQVKNLVHVPKIQRGPKHLKVQVGQORVIDPCNAOQTPLP 1200
QY 127 -----GE-----N-----T-----A-----N-----G----- 133
Db 1201 VITWSKGSMTLVDGEHVSNDPGTSLSDOATPSDAGIYTCVATNIAGTDETEITLHVQE 1260
QY 134 -----E-----E-----N-----GA-----H----- 139
Db 1261 PPTVEDELPYNTTFOERVANQRIEFPCKAGTPKPTIKWHLNGLRELTGREPGISILEEG 1320
QY 140 T---IA-----N-----N-----HT-----DM----- 148
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKQEVNSVLLNQLTNL 1380
QY 149 -----M-----E----- 150
Db 1381 FCEVEGTSPPIIMWYKDNQVVBESSTIQTVNNKGILKLFATPEDAGRYSCKAINIAGTS 1440

QY 151 -----VD-----G-----DV-----E-----I-----P-----PN 160
Db 1441 QKYFNIDVLVPPTIIGTNEPKEVSVVLNRDVALEQCVKGTPFPDIHWFKDGKFLFLGDPN 1500
QY 161 -----K-----A-----V----- 165
Db 1501 VELLDRGQVLHLKARNDKGRYQCTVSNAAAGQAKDIKLTIVIPPSIKGNVTTDISVL 1560
QY 166 -----RG-----H----- 168
Db 1561 INSILKLECKTGLPMPAITWYKQOPIMSSSOALYIDKQYLHI PRAQVSDSATYTCV 1620
QY 169 -----E-----S-----E-----VFI-----C-----A-----W----- 177
Db 1621 ANVAGTAEKSFHVVVVPPMIEGNLATPLNKQVIAHSLTLECNAAAGNPSILTLKDG 1680
QY 178 -----N-----P-----V-----SD----- 182
Db 1681 PVKANDNFRIEAGKKLEIMSAQIEDRGYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740
QY 183 -----LL-----ASGS-----G-----D-----S----- 191
Db 1741 TSYFIVWNNLLELDCHVTGSPPTIMLWKDQQLIDERDGFKILLNGRKLVIQAQVSN 1800
QY 192 -----TA-----R-----I-----W 196
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSSGLSERVWVKYKPVALQCIANGIPNPSITW 1860
QY 197 -----NL-----S----- 199
Db 1861 LKDDQPVNTAQNLIKQSSGRVLQIAKTLELDAGRYTCVATNAAGETQOHIQHLVHPPS 1920
QY 200 -E-----N-----S-----T-----SGST----- 207
Db 1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLLSSGSTMTFLNRQIIDIES 1980
QY 208 -Q-----L-----V-----LR-----H-----C-----I----- 215
Db 1981 AQISDAGIVKCAINSAGATELFYSLQVHVAPSISGNNMVAVVVNNVRLECEARGIPA 2040
QY 216 -----R-----E-----G-----G-----O-----D----- 221
Db 2041 PSLTWLKDGPVSSFSNGLQVLSGGRILALTSTQISDTRGYTCVAVNAAGEKORDILRV 2100
QY 222 -VPSN-----KD-----VT-----S-----L----- 231
Db 2101 YVPPNIMGEQNVSVLISOAVELLQSDAIPPTTLTWLKDGHPLKKPGLSISENSVLK 2160
QY 232 -----D-----W-----N-----S-----EG-----TLL----- 240
Db 2161 IEDAQVQDTRGYTCEATNVAGKTEKKNYNNIWPVPPNIGSDELTQLTWIEGNLSLCE 2220
QY 241 AT-----GS-----Y-----D-----G-----F-----A----- 249
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRLISGGRQLOISIAEKSDAALYSCVASNVAGTAK 2280
QY 250 -----R-----R-----I-----WTKDG----- 256
Db 2281 KEYNLQVYIRPTITNSGSHPTBIIVTRGKSISLECEVQIIPPTVTWMDGHPHLLIKAGV 2340
QY 257 -----NL-----AS-----T-----L-----GOH----- 265
Db 2341 EILDGHIQLQKNIHVSDTRGYVCVAVNVAGTDDKYDLSVHAPPSIIGNHRSPIENVV 2400
QY 266 -K-----G-----PI-----FALK-----W----- 274
Db 2401 EKNSVLTCEASGIPLPSTWF--KDGWPVLSNSVRILSGGRMLRLMQTTMEDAQOYTC 2458
QY 275 -----N-----K-----K-----GN----- 279
Db 2459 VVRNAAGEBERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518
QY 280 -----F-----I-----L-----S-----AGV----- 286

Db 2519 QPLQDEAHHTISGRFLQITNVQVPHTRVTCCLASSPAGHKSRSFSLNVFVSTIAGVG 2578
QY 287 D--K-T-----T-I-----I-W----- 293
Db 2579 SDGNPEDVTILNSFTSLVCBAYSYPATITWFKDGTPLSNRNIRILPGORTLOILNAQ 2638
QY 294 D-----A-----H-----TG-----E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVYIPPIINKGDLWGPGLSPKEVKIKVNNLTLECE 2698
QY 300 A-----K-----Q----- 303
Db 2699 AYATPSASLWYKDGQPLKSDHVNIAANGHTLQI KEAQISDTGRTYTCVASNIAGEDEL 2758
QY 304 F-----P-F-----H----- 307
Db 2759 FDNVNIQPPSFQKLWEIGNMLDTRNGEAKDVIIINNPISLYCETNAAPPPTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTSSDKVLIILPGGRVLQIPRAKVEDAGRYTCVAVNEAGESLSQYDVRVLVPIIEGANS 2878
QY 316 -----WQ-----SN----- 319
Db 2879 LPEBVTLVNKSALIECLSSGSPAPRNSWQKGOPLLEDHDKFLSNGRILQILNTQITD 2938
QY 320 -----NT-----PAS-C-----S 326
Db 2939 IGRYVCAENTAGS AKYFNLNVHVPSPVIGPKSENLTVVVNNFISLTCEVSGFPDPLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKXQPIKLTNTLIVPGERTLQIIRAKVSDGGEYTCIAINXAGESKKFSTVTVVPPS 3058
QY 332 --H-----V-C-----K-----L-GQ----- 338
Db 3059 IKDHSESLVMNVREGTSVLECSNAVPPPVITWYKNGRMITESTHVEILADGQMLHI 3118
QY 339 -----D-----R-----P-IX----- 343
Db 3119 KKAESVDTGVYCRAINVAGRDKNFNLNVVYPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 --T-----F-----Q-----G-----H-----T----- 349
Db 3179 IPPTIAWLKXKRIENSLSLEVRILSGGKLQIARSOHSDSGNYTCIASNMEGKAQYY 3238
QY 350 -----N-E-V-NA-----IKW-D-P-----T----- 360
Db 3239 FLSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIQMLKDKGPIASGETERIRV 3298
QY 361 --G--NLL-A-S-----C-S-----D-----D-----MTL-- 373
Db 3299 SANGSTLNIYGALTSDTGKYTCVATNPAGBEDRIFNLNVYVPTIRGNKDEAEKLMYVD 3358
QY 374 --K-----I-W-----SM-----KQ-D-----NCV-- 384
Db 3359 TSINIECXTGTPPPQINWLKNGLPPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVSN 3418
QY 385 -----H-----Q-----D-----L-Q----- 388
Db 3419 RAGVDNKHVNLQVAFPPNMDNSMGTERITVLKGSSTSMACITDGTTPAPSMAWLRDQPLG 3478
QY 389 -----Q-----H-N-----KE 393
Db 3479 LDAHLTVSTHGMVQLLKAEETDSGKYTCIASNEAGEVSKHFLKVLPEPPHINGSBEHE 3538
QY 394 I-----YT-----I-K-W----- 399
Db 3539 ISVIVNNPLELTCIASGIPAKMTWMDKGRPLQTDQVOTLGGGEVLRISTAQVEDTRY 3598
QY 400 ---SPTG-----P--GT-----N----- 407

Db 3599 TCLASSPAGDDKXYLVRVHVHPNPIAGTDEPRDITVLNRQVTLCKSDAVPPPVITWL 3658
QY 408 N-----P-----N-----ANUM-LAS-A-----S-F----- 420
Db 3659 NGERLOATPRVRLISGRVLIQINNADLGDNTANTCVASNIAGTKTREFILTVNVPPIK 3718
QY 421 -----DSTV-----RL-W-D-V-----DR-----G--I----- 433
Db 3719 GPSSLVILLNKSTVLECIAGVPTPRITWRKOGAVLAGNHARYSILENGFLHIQSAHVTD 3778
QY 434 ---C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCWATNAAGTDRRIDLVHVPSPSIAGPTNTMTVIVNVQVTLTACEATGPKPSINW 3838
QY 440 -K-H-----Q-E-----P-V-----Y----- 446
Db 3839 RKNGHLLNVQNSYRLLSSGSLVITSPSVDTTATYECTVTNGAGDKRTVLTVDVQVPP 3898
QY 447 SVA-----F-S-----P-D----- 453
Db 3899 SIADEPTDFLVTKHAPAVITCTASGVPPFSIHWTKNGIRLLPRGDGYRILSSGAIEILAT 3958
QY 454 ---GRY-----L-----ASG-S- 461
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLHVHBPVQPOPSSELHVILNNPILLPCEATGTPSP 4018
QY 462 F-----D-----KCV----- 466
Db 4019 FITWQKEGINVNTSGRHNHVLPSGGLQISRAVREDAGTYMCVAQNPAQTALGKIKLVQV 4078
QY 467 ---H-----I-W-----NT-Q-- 472
Db 4079 PPVISPHLKEYVIAVDKPIITLSCAADGLPPDPITWHKDGRAIVESTIRQVLSGSLQIAF 4138
QY 473 V-----C-----L-----HY----- 477
Db 4139 VQPGDAGHYTCMAANVAGSSSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGPT 4198
QY 478 ---L-N-----GO-----VL-----L-----N----- 485
Db 4199 PAINKKONVLLANLLGKYTABPYGELILENVVLEDSGYFTCVANNAAGEDTHVSLTVH 4258
QY 486 ---L-G--R-SIC-----L-YT-----LP-H-----H-LV-- 500
Db 4259 VLPTTTELPGDVLNKGEOQLRLS-CXATGIPLPKLTWTNNNIIPAHFDSVNGHSELVIE 4317
QY 501 -----V-----I-PL-----V-- 505
Db 4318 RVSKEDSGTYVCTAENSUGFVKATGFVYVKEPPVFKGDPYPSNWIPLGNAILLNCEVKGD 4377
QY 506 ---AL-----IE-----L-----L-VL-----K 514
Db 4378 PTFTIOWNRKGVDDIEISHRIRQLNGSLAIYGTVNE 4413
RESULT 15
ID ADJ83137
XX ADJ83137 standard; protein; 5636 AA.
XX ADJ83137;
XX 06-MAY-2004 (first entry)
DE Human hemimentin protein - SEQ ID 128.
XX NOVX; cytostatic; antiarteriosclerotic; antidiabetic; antiasthmatic;
KW antiallergic; antiinflammatory; respiratory; antiarthritic;
KW dermatologic; antibacterial; cerebroprotective; vasotropic; cardiac;
KW haemostatic; hypotensive; hepatotropic; neuroprotective; anorectic;
KW neoplastic; antitumor; muscular; immunosuppressive; gynaecological;
KW antiparkinsonian; anticonvulsant; tranquiliser; analgesic; nephrotropic;
KW antiinfertility; antilipaemic; cardiomyopathy; atherosclerosis; diabetes;

cell signal processing; metabolic pathway; asthma; allergy; emphysema;
autoimmune; graft-versus-host; arthritis; cancer; stroke; haemophilia;
obesity; Alzheimer's; pain; chromosome mapping; tissue typing; human;
hemocentin.

Homo sapiens.

US2003170630-A1.

11-SEP-2003.

21-DEC-2001; 2001US-00032189.

21-DEC-2000; 2000US-0257495P.

22-DEC-2000; 2000US-0258171P.

20-FEB-2001; 2001US-0269940P.

08-MAR-2001; 2001US-0274192P.

22-MAR-2001; 2001US-0277826P.

29-MAR-2001; 2001US-0279840P.

11-APR-2001; 2001US-0282981P.

13-APR-2001; 2001US-0283656P.

31-JUL-2001; 2001US-0309247P.

10-AUG-2001; 2001US-0311754P.

17-AUG-2001; 2001US-0313331P.

(ALSO/) ALSOBROOK J P.

(TCHE/) TCHERNEV V T.

(LIUX/) LIU X.

(SPVT/) SPYTEK K A.

(ZERH/) ZERHUSEN B D.

(PATT/) PATTURAJAN M.

(LEPL/) LEPLY D M.

(BURG/) BURGESS C E.

(SHIM/) SHIMKETS R A.

(GROS/) GROSSE W M.

(SZEK/) SZEKERES E S.

(VERN/) VERNET C A M.

(LILL/) LI L.

(CASM/) CASMAN S J.

(BOLD/) BOLDOG F L.

(GORM/) GORMAN L.

(GANG/) GANGOLLI E A.

(FERN/) FERNANDES E R.

(RIEG/) RIEGER D K.

(EDIN/) EDINGER S R.

(GUNT/) GUNTHER E.

(MILL/) MILLET I.

(SCIO/) SCIORE P.

(ELLE/) ELLERMAN K.

(MACD/) MACDOUGALL J R.

(SMIT/) SMITHSON G.

Alsobrook JP, Tchernev VT, Liu X, Spytek KA, Zerhusen BD;

Patturajan M, Lepley DM, Burgess CE, Shimkets RA, Grosse WM;

Szekeres EA, Vernet CAM, Li L, Casman SJ, Boldog FL, Gorman L;

Gangolli ES, Fernandes ER, Rieger DK, Edinger SR, Gunther E;

Millet I, Sciore P, Ellerman K, Macdougall JR, Smithson G;

WPI; 2003-898249/82.

New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,
preventing or treating NOVX-associated polypeptide disorder, e.g.
cardiomyopathy, atherosclerosis, diabetes, cancer, Parkinson's disease or
asthma.

Disclosure; SEQ ID NO 128; 263pp; English.

The invention relates to a novel isolated NOVX polypeptide. The
polypeptide demonstrates cytostatic, antiarteriosclerotic, antidiabetic,
antiasthmatic, antiallergic, antiinflammatory, respiratory,
antiarthritic, dermatological, antibacterial, cerebroprotective,
vasotrophic, cardiant, haemostatic, hypotensive, hepatotropic,
neuroprotective, anorectic, nootropic, antilucer, muscular,

CC immunosuppressive, gynaecological, antipsoriatic, endocrine,
ophthalmological, osteopathic, antiparkinsonian, anticonvulsant,
CC tranquilliser, analgesic, nephrotropic, antiinfertility and antilipaemic
CC activities. The NOVX polypeptide, nucleic acid or antibody of the
CC invention may be useful for treating or preventing a NOVX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, diabetes or a disorder
CC related to cell signal processing and metabolic pathway modulation.
CC Furthermore, the NOVX polypeptides may be useful for diagnosing, treating
CC or preventing diseases such as asthma, allergies, emphysema, autoimmune
CC disease, graft-versus-host disease, arthritis, cancer, stroke,
CC haemophilia, obesity, Alzheimer's disease and pain. The nucleic acids may
CC be used as hybridisation probes, in chromosome mapping, tissue typing,
CC preventive medicine or pharmacogenomics. The current sequence is that of
CC a protein of the invention which is related to human NOVX protein.

XX SQ Sequence 5636 AA;

Query Match 71.2%; Score 2596.5; DB 7; Length 5636;

Best Local Similarity 9.7%; Pred. No. 1.1e-30;

Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

QY 1 M-S-----I-----SS---D-----EV----- 8

Db 1 MISWEVVHTVFLFALLYSLAODASPOSEIRAEPEEGASTLAFVFDVTCGMYDDLQVVI 60

QY 9 -----NF-LV----- 12

Db 61 EGASKILETSLKRPKPLFNFLVPHDPFPGVTTTDPKKFQYELRELIVGGGDCPE 120

QY 13 -----Y---R---Y---LQ----- 17

Db 121 MSIGAIIKALEISLPGSFYVFTDARSKDYRLTHEVLTIQKQSVFVLTDGCDRTH 180

QY 18 -----E-----SG--F-----S-H--S-----A-----F 26

Db 181 IGYVVEEIASTSSGQVFLDKKQVNEVLKVEEAVQASKVHLLTDHLEQAVNTWRIPF 240

QY 27 -----T-----FG-----I-----K 31

Db 241 DPSLKEVTVLSGSPSMIEIRNPLGKLIKKGFLHELLNIHNSAKVNVNKEPEAGMTVK 300

QY 32 ---S-H-----I-----S---Q----- 36

Db 301 TSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSQVGIPTVYLLNTSGISTPARID 360

QY 37 -----S-----NI-----N-----G----- 41

Db 361 LLELISIGSSSLKTIPTVKYYPHRKPYGIWNI SDFVPPNEAFFLKVTGYDKDYLFORVSS 420

QY 42 -----A-----L-----VP-----P-----A 47

Db 421 VSFSSIVDPAPKVTMPKTPGYVLPQGI PCSVDSLPLPTLSFVRNGVTLGVDQYLKESA 480

QY 48 -----A---L-----I-----S---IIQ-----KG----- 56

Db 481 SVSLDLAKVTLSDEGEYECIAVSSAGTGRAQTFDFVSEPPVQVPPNVTVTPGERAVLT 540

QY 57 -----L--Q-----L-----Y---V---E----- 61

Db 541 CIIISAVDNLTWQRNDRVRLAEPARIRTLANLSLELKSVPKNDAGEYHCVMSSEGGSS 600

QY 62 -A-----EVS1-----NE----- 68

Db 601 AASVFLTVQEPKVTVMKQSQFTGSGSEVSI MCSATGYPKPKIATWNTDMFIVGSHRYRM 660

QY 69 --DGLTF-----D-----G-----R-----P-----I-- 78

Db 661 TSDGTLFIKNAAPKDAIGYCLASNSAGTDKQNSTLRYIEAPKLMVQSELLVALGDITV 720

QY 79 -E---S-----L--S--LI-----DA-----V----- 87

Db 721 MECKTSGIPPPQVKNFKGDLRLRPSTFLIIDPLLLGLLK1QETODLDAGDYTCVAINAAGR 780

QY 88 -----M-----P-DV-----VO-----T-----94
Db 781 ATGKITLDVGPVPFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMFI SRPFS 840
QY 95 -----R-----95
Db 841 VSSISQLRTGALFILNLWASDKGTICAEANQFGIKQSETTVTVTGLVAPLIGISPSVAN 900
QY 96 ---OO---A---Y---R---D---KL---103
Db 901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLLQNPYITVRS DGLSHIERVLOQDGGEYTC 960
QY 104 ---A---Q---O---AA---109
Db 961 VASVAGTNNKTTSVVHVLPITIOHQOILLSTIEGIPVTLPCASGNPKPSVIMSKKGL 1020
QY 110 ---AAA---A---AA---A---118
Db 1021 ISTSAKFSAGADGLYVWSPEGESEYVCTATNTAGYAKRVQLTVYVPRVFGDLRG 1080
QY 119 -SQ-----Q-----123
Db 1081 LSQDKPVEISVLAGEVTLPCFVKSLPPIITWAKETQLISPPSRHTPLPSSGSMKITET 1140
QY 124 ---A---K---N---126
Db 1141 RTSDSGMYLCVATNIAGNVTAQVKNLVHPKIQRGPKHLKVQVGQVDPICNAQGTPLP 1200
QY 127 ---GE---N---T---A---N---G---133
Db 1201 VITWSKGGSTMLVDGHHVSNPDGTLISIDQATPSDAGIYCVATNIAGTDETEITLHVQE 1260
QY 134 ---E---E---N---GA---H---139
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEFPCKAGTKPTIKWLHNGREL TGREPGISILEEG 1320
QY 140 T---IA---N---N---HT---DM---148
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKQEQVSNVLLNQLTNL 1380
QY 149 ---M-----E-----150
Db 1381 FCEVEGTPSPPIIMWKDNQVQTESSTIQTVNGKILKLFATPEDAGRYSCAKAINIAGTS 1440
QY 151 ---VD---G---DV---E---I---P---PN---160
Db 1441 QKYFNIDLVPPPTIIGTFNFKVSVLNRDVALEQCQVKGTPFPDIHWFQDKGLFLGDPN 1500
QY 161 -----K---A---V---VL---165
Db 1501 VELLDRGOVLHKNARNDKGRYQCTVNSAAGQAQDKLTIYIPPSIKGNNVTDISVL 1560
QY 166 ---RG-----H---168
Db 1561 INSLIKLECKTRGLPMAITWYKDGQFIMSSQALYIDKQYLHI PRAQVSDSATYTCV 1620
QY 169 ---E---S---E---VFI---C---A---W---177
Db 1621 ANVAGTAKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECNAGNPSPIITWLDGV 1680
QY 178 ---N-----P---V---SD---182
Db 1681 PVKANDNFRIEAGGKKLEIMSAQEDIRGQYICVATSVAGEKEIKYEDVDVLVPPAIEGGDE 1740
QY 183 ---LL---ASGS---G---D---S---191
Db 1741 TSYFIWVNNLLELDCHVTGSPPTIMWLKDGQLIDBERDGFKILLNGRKLVI AQAQVSNT 1800
QY 192 ---TA---R-----I---W---196
Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSSGSLSERVVVKYKPVALQCIANGIPNPSITW 1860
QY 197 ---NL-----S---199

Db 1861 LKDDQPWTAQGNLKIQSSGRVLQIAKTLLEDDAGRYTCVATNAAGETOQHQLHVHPPS 1920
QY 200 -E---N---S---T---SGST---207
Db 1921 LEDAGKMLNETVLVSNPVLQECKAAGNPVPVITWYKDNCLLSGTSMTFLNRGQIIDIES 1980
QY 208 -O---L---V---LR-H---C---I---215
Db 1981 AQISDAGIYKCAVINSAGATELFYSLQVHVAPSISGNNMNVAVVANNPVRLECEARGIPA 2040
QY 216 ---R---E---G---G---O---D---221
Db 2041 PSITWLDKGSVPSSPENGLOVLSGGRILALTSTQISDTGRYTCVAVNAAGEKORDILRV 2100
QY 222 -VPSN---KD---VT---S---L---231
Db 2101 YVPPNIMGEBQNVSVLISOAVELLCSQDAIPPTLTWLDKGHPLLKPKGLSISENRSVLK 2160
QY 232 ---D---N---S---EG---TL---240
Db 2161 IEDAQVODTGRYTCBATNVAGTEKKNYNNIHWPPNIGGSDDEL TQITVIEGNLISLCE 2220
QY 241 AT---GS---Y---D---G---F---A---249
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRLISGGRQLQISIAEKSDAALYSCVASNVAGTAK 2280
QY 250 ---R---I---WTKDQ---256
Db 2281 KEYNLQVIRPTITNSGSHPTETIIVTRGKSISLECEVOGIPPTVTWMDKGHP LIAKAGV 2340
QY 257 ---NL---AS---T---L---GOH---265
Db 2341 EILDEGHILQKNHIVSDTGRYVCVAVNAGWMDKKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K---G---PI---FALK---W---274
Db 2401 EKMSVSLTCEASGIPLPSTTWP--KDGWPVLSLSNVRLISGGRMLRMQTTMEDAGQYTC 2458
QY 275 ---N---K---K---GN---279
Db 2459 VVRNAAGEERKIFGLSVLPVPHIVGENTLEDVKVEKQSVTLTCEVTGNPVPBITWHKDG 2518
QY 280 ---F---I---L---S---AGV---286
Db 2519 QPLODEAHIIISGGRFLQITNVQVPHYTCCLASSPAGHKRSFSLNVFVSPPTIAGVG 2578
QY 287 -D---K---T---T---I---I---W---293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPPATITWPKDGTPLESNRNIRILPGGRTLQILNAQ 2638
QY 294 -D---A---H---TG---E---299
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKYVYIPPI INKGLWPGGLSPKEVKIKVNNLTLECE 2698
QY 300 A---K---Q---O---303
Db 2699 AYAI PSASISLWYKDGQPLASDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL D 2758
QY 304 F---P---P---LDV---H---307
Db 2759 FDVNIQVPPSPFQKWEIIGNMLDTGRNGEAKVDIINNPI SLYCETNAAPPPTLTWYKDGHP 2818
QY 308 --SA---P---A---LDV---D---315
Db 2819 LTSSDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS D 2878
QY 316 ---WQ---SN---319
Db 2879 LPEEVTVLVKNKALIECLSSGSPAPNSWQKQDQPLEDDHKKFLSNGRILQILNTQITD 2938
QY 320 ---NT---FAS---C---S---326

Db 2939 IGRYVCAENTAGSAXKYFNLNVHVPSPVIGPKSENLTVVNNFISLTCEVSGFPPDLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKXQPIKLTWTLVPGGRTLOIRAKVSDGGEYCTAINXAGSXXXKFSITVVVPS 3058
QY 332 H-----V-----C-----K-----L-GQ----- 338
Db 3059 IKDHSESLSVNVREGTSVLECESNAVPPVITWYKGRMITESTHVEILADGQMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KKAESDTGOYVCRAINVAGRDDKNFLNLYVPPSIEGPEREVITISNPVTLTCDATG 3178
QY 344 -----T-----F-----Q-----G-----H-----T----- 349
Db 3179 IPPPTTAMLKXKYKRIENSLSVRILSGGSKLQIARSQSDSGNYCTIASNMEGKAQYY 3238
QY 350 -----N-E-V-NA-----IKW-D-P-----T----- 360
Db 3239 FLSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIQWLDGKPIASGETERIRV 3298
QY 361 -G--NLL-A-S-----C-S-----D-----D-----MTL-- 373
Db 3299 SANGSTLNIYALTSDGKYTCVATNPAGEDRIFNLNVVYPTIRGNKDEABKLTMYD 3358
QY 374 -----K-----I-W-----SM-----KQ-D-D-----NCV-- 384
Db 3359 TSINIECRXTGTPPPQINWMLKNGPLPLSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418
QY 385 -----H-----Q-----D-----L-----Q----- 388
Db 3419 RAGVDNKHYNLQVFAFPNDMSMGTEBITVLKGSSTSMACITDGTAPSMWLDRDQPLG 3478
QY 389 -----O-----H-N-----KE 393
Db 3479 LDAHLTVSTHGMVQLLKAETEDSGKYTCIASNEAGEVSKHFLKVLPEPHNGSEHEE 3538
QY 394 I-----YT-----I-K-W-----N----- 399
Db 3539 ISVIVNPLELTCTIASGIPAPKMTWMDGRPLPQTDQVTLGGGEVLRISTAQVEDTRY 3598
QY 400 -----SPTG-----P-----GT-----N----- 407
Db 3599 TCLASSPAGDDDKYELVRVHVHPNIAGTDEPRDITVLRNQVLTCKSDAVPPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS-A-----S-F----- 420
Db 3659 NGERLOATPRVIRLSGGRYLQINNADLGDYANTYTCVASNIAGKTTREFILTVNVPPNIKG 3718
QY 421 -----DSTV-----RL-W-D-V-----DR-----G-I----- 433
Db 3719 GPOSVLILLNKSTVLECIAGVPTPRITWRKOGAVLAGNHARYSILENGFLHIQSAHVTD 3778
QY 434 -----C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCMATNAAGTDRRLDQVHVPPSIAPGPTNMTVIVNQVTLTACEATGIPKPSINW 3838
QY 440 -K-H-----Q-E-----P-V-----Y----- 446
Db 3839 RKNHLLANDQNSYRLLSGSLVLIISPSVDDTATYECTVTNGAGDKRKTVDLTQVPP 3898
QY 447 SVA-----F-S-----P-D----- 453
Db 3899 STADEPTDFLVTKHAPAVITCTASGVPPFPIHWTKNIGIRLLPRGDGYRILSSGAIELAT 3958
QY 454 -----GRY-----L-----ASG--S- 461
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIQPOPSELHVILNNPILLPCEATGTPSP 4018
QY 462 F-----D-----KCV----- 466
Db 4019 FITWQKGINVNTSGRHAFLPSGGQLISRAVREDAGTYMCVAQNPAQTALGKILNVQV 4078

QY 467 -----H-----I-W-----NT-Q--- 472
Db 4079 PPVISPHLKEYVIADVDPITLSCADGLPPPDITMHKDGRAIVESIRQVLSGSLQIAF 4138
QY 473 V-----C-----L-----HY----- 477
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGIPT 4198
QY 478 -----L-N-----GO-----VL-----L-----N----- 485
Db 4199 PAINWKKONVLLANLLGKYTAEPYGEILLENVVLEDSGFYTCVANNAAAGEDTHVSLTVH 4258
QY 486 -----L-G-----R-SIC-----L-YT-----LP-H-----H-LV-- 500
Db 4259 VLPTFTFELPGDVSLNKGEOQLRLS-CKATGILPLPKLTWTFNNNIIPAHFDSVNGHSELVIE 4317
QY 501 -----V-----I-PL-----V----- 505
Db 4318 RVSKEDSGTYVCTAENSVGFKATGFVYVKEPPVFKGDYPSNWIPLGGNAILNCEYKGD 4377
QY 506 -----AL-----IE-----L-----VL-----K 514
Db 4378 PTFTIOWNRKGVDBIISHRIRQLGNSGLAIYGTVNE 4413

Search completed: January 3, 2005, 15:19:58
Job time : 116.333 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:10:17 ; Search time 23.6667 Seconds
(without alignments)
1440.315 Million cell updates/sec

Title: US-09-987-701-4
Perfect score: 3649
Sequence: 1 MSISDEVNFLVRYLQESG.....LPHLVVIVLALIELVLK 514

Scoring table: BLOSUM30
Gapop 1.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pdp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pdp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pdp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pdp.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pdp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2571.1	70.5	3461	US-09-334-220-2	Sequence 2, Appli
2	2562.6	70.2	4544	US-08-469-486-52	Sequence 52, Appl
3	2562.6	70.2	4544	US-08-469-658-52	Sequence 52, Appl
4	2558.9	70.1	4654	US-08-476-515A-84	Sequence 84, Appl
5	2558.8	70.1	4655	US-08-652-877-84	Sequence 84, Appl
6	2558.8	70.1	4655	US-08-652-877-88	Sequence 88, Appl
7	2558.8	70.1	4655	US-08-652-877-90	Sequence 90, Appl
8	2557.8	70.1	4655	US-08-652-877-86	Sequence 86, Appl
9	2554	70.0	3460	US-09-334-220-1	Sequence 1, Appli
10	2552.2	69.9	5405	US-08-718-388-9	Sequence 9, Appli
11	2533.4	69.4	15281	US-08-471-119A-2	Sequence 2, Appli
12	2527.7	69.3	4861	US-09-919-497-70	Sequence 70, Appl
13	2527.4	69.3	4303	US-08-460-751-2	Sequence 2, Appli
14	2527.4	69.3	4303	US-09-479-467A-2	Sequence 2, Appli
15	2526.5	69.2	4302	US-09-052-469-8	Sequence 8, Appli
16	2526.5	69.2	4302	US-08-422-582-8	Sequence 8, Appli
17	2526.5	69.2	4302	US-09-052-262-8	Sequence 8, Appli
18	2526.5	69.2	4339	US-09-052-469-6	Sequence 6, Appli
19	2526.5	69.2	4339	US-08-422-582-6	Sequence 6, Appli
20	2526.5	69.2	4339	US-09-052-262-6	Sequence 6, Appli
21	2523.5	69.2	3571	US-09-911-842A-2	Sequence 2, Appli
22	2511.4	68.8	4302	US-08-658-136-5	Sequence 5, Appli
23	2477.3	67.9	3594	US-09-911-842A-4	Sequence 4, Appli
24	2463.6	67.5	4866	US-08-424-783-2	Sequence 2, Appli
25	2460.1	67.4	3727	US-08-222-617A-2	Sequence 27, Appl
26	2459.4	67.4	3959	US-08-970-269A-30	Sequence 30, Appl

28	2459.4	67.4	3959	3	US-09-407-562-30	Sequence 30, Appl
29	2459.1	67.4	3666	2	US-08-222-617A-12	Sequence 12, Appl
30	2456	67.3	4928	3	US-09-036-987A-5	Sequence 5, Appli
31	2456	67.3	4928	3	US-09-370-700-5	Sequence 5, Appli
32	2456	67.3	4928	4	US-09-603-207-5	Sequence 5, Appli
33	2451.4	67.2	7257	3	US-09-335-409-5	Sequence 5, Appli
34	2451.4	67.2	7257	3	US-09-568-102-5	Sequence 5, Appli
35	2451.4	67.2	7257	3	US-09-567-969-5	Sequence 5, Appli
36	2451.4	67.2	7257	3	US-09-568-480-5	Sequence 5, Appli
37	2451.4	67.2	7257	3	US-09-568-486-5	Sequence 5, Appli
38	2451.4	67.2	7257	3	US-09-568-472-5	Sequence 5, Appli
39	2451.4	67.2	7257	3	US-09-567-899-5	Sequence 5, Appli
40	2434.6	66.7	4536	4	US-09-079-030-1	Sequence 1, Appli
41	2434.6	66.7	4563	4	US-09-538-092-842	Sequence 842, App
42	2433.9	66.7	5032	4	US-09-538-092-979	Sequence 979, App
43	2433.6	66.7	4563	4	US-09-108-006C-1	Sequence 1, Appli
44	2433.5	66.7	3898	3	US-08-750-717-2	Sequence 2, Appli
45	2432.6	66.7	4536	4	US-09-180-422B-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-09-334-220-2
; Sequence 2, Application US/09334220
; Patent No. 6323177
; GENERAL INFORMATION:
; APPLICANT: St. Jude Children's Research Hospital
; APPLICANT: Curran, Thomas
; APPLICANT: D'Arcangelo, Gabriella
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND
; TITLE OF INVENTION: THERAPIES
; FILE REFERENCE: 2427/0F704
; CURRENT APPLICATION NUMBER: US/09/334,220
; CURRENT FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 3461
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-334-220-2

Query Match	70.5%	Score	2571.1;	DB	3;	Length	3461;
Best Local Similarity	11.5%	Pred. No.	6e+33;				
Matches	396;	Conservative	94;	Mismatches	16;	Indels	2939;
Gaps	345;						
QY	1	M	-----SI-----S-----S-----S-----D-----EV-----8				
Db	1	M	-----SI-----S-----S-----S-----D-----EV-----8				
QY	9	N	-----N-----P-----LV-----YRY-----LQ-----ES-----GF-----S-H 23				
Db	61	H	AGNPTVYPGQYHVTISTSTFFDGLLVTLG-YTSTSIQSSQSSGSAFGGMSDH 119				
QY	24	S	-----S-----A-----P-----T-----P-----G-----I-----K-----31				
Db	120	Q	FGNQPMCSVWASHVSHLPTNLSPVYIAPPAGTCVNFPMATATHRGOVIFKDALAQOLC 179				
QY	32	S	-----SH-----I-----S-----Q-----SNI-----N-----GAL-----43				
Db	180	E	QAPTEATAYSHLAIHSDSVILRRDFFSYQQLNPNLWVCSNCEMGECQGTINMGN 239				
QY	44	V	-----P-----P-----A-----AL-----I-----SI-----IQ-----54				
Db	240	A	VTFCPEYGPRLTTTCLNTTASVLFQFSIGSGSCRFSDPSITVYAKNNTADMTQLE 299				
QY	55	K	-----KG-----LQY-----V-----EA-----E-----VSIN-----67				
Db	300	K	RAPSNVSTIHLPLPBEAKGESVQFQWQDLSRVGEVYACWALDNLV-INSahre 358				

QY 68 ---ED---GT-LF---DG---R---76
Db 359 VVLENDLPVDTGNWLFPGATVKHSCQSDGNSIYFHGNEGSEFNATTRDVLSTEDIQ 418
QY 77 ---P---I---ES-LSLI---DA---86
Db 419 EQWSEFESQPTGWDILGAVVGADCGTVESGLSVFLKDGCKLCTPYMDTTGYGNLRFY 478
QY 87 -VM---P---D---VW---QT---94
Db 479 FVMGGICDPCGVSHENDIILYAKIEGRKEHIALDTLTYSYKVPSPVSVINPELQTPATK 538
QY 95 ---RO---QAY-R---D---K-LAQ---Q---106
Db 539 FCLURQXSHQYRNVWAVDFHVLPLVLPSTMSHMIQPSINLGCCTHQPNGNSVSLEFSTNH 598
QY 107 -QA---A---AAA---AA---114
Db 599 GRWSLLHTECLPEICAGPHLPSTVYSSYSGMNRITITPLPNAALTRDTRIRWRQTGP 658
QY 115 ---A---A---AASQ---Q-GS 123
Db 659 ILGNMWAIDNVYIGPCLKFCGSGQCTRHGCKDPGSGPACEMASQTFPMFISESFGS 718
QY 124 AK---N---G-E---N---129
Db 719 ARLSSVHNFYIRGAEVSGCGVLASGKALVFNKDGRRQLITSPLDSSQSRFLQFTLRIG 778
QY 130 ---T-A---NG---E-E---N---G-137
Db 779 SKSVLSTCRAPDQGEVLLHYSDNGITWKLLEHYSYVNYHPRISVLPDDARQFGI 838
QY 138 ---A-H---T---I---AN---N---H-145
Db 839 QFRWQPYHSSQGEDVWAIDEIVMTSVLFNSISLDEFTNLVEVTQSLGFLGNVQPYCGHD 898
QY 146 ---T-D---M---NEVDG---D-V-E---156
Db 899 WTLCTGDSKLASMRVETQSMQI-GASYMIQFSLVMGCGQKYTHMDNQVKLEYSANH 957
QY 157 ---I---P-PN-K---A---VVL---R---166
Db 958 GLTWHLVQBECLSPMSCQFTSASIVHASEFTQWRRTVVLQKTSWGATRFWMSQYY 1017
QY 167 ---GH---E-SE---V-P---173
Db 1018 TAQDEWALDNIYIGQCPNMCSGHGCDHGVCRCDOGYQTECHPEAALPSTIMSDFENP 1077
QY 174 ---I---C---A---W---177
Db 1078 SSWESDQEVIGEVVKPQCGVSVSGSSLYFSKAGKROLVSWDLDTSWDFVQFYQI 1137
QY 178 ---N-P---V---SD---L-L-ASG---187
Db 1138 GGESAACNKPDSREEGILLQYSNNGGIQWHLAEMFSDFSKPRFVYLELPAAGKTPCTR 1197
QY 188 ---SG---D---S---T---A---R-IW-196
Db 1198 FRWKPVSGEDYDQWAVDDIILSEKQKQVIVPNVPTLPQNFYKPAFDYPNMQMSVWL 1257
QY 197 ---N-LSEN-S---T---S---G---ST 207
Db 1258 MLANEGMAKNDSCATPFSANVFGKSDGRPAVTRDTLTKPGVYLQPKLNGICTSPSST 1317
QY 208 ---Q---L---V---210
Db 1318 APVLQYSHDAGMSWFLKEGCFPASAAGCEGNSRELSPTVYVYTGDPBEWTRITIAIP 1377
QY 211 ---L---R---H-CI---215
Db 1378 RSLASSKTRFRWIOESSQKNVPPFGIDGYVISEPCPSYCSGHGDCISGVCFDLGYTAA 1437
QY 216 ---R-E---GGQ---D---V 222

Db 1438 QGTCVSTNPHSEMDFRFBGKLSPLWYKITGQVGTGCTLNDGRSLYFNLGKREARTV 1497
QY 223 P---SN---K-D---227
Db 1498 PLDTRNISLVQFYIQIGSKTSGITYITPRARYEGLVVQVSYNDNGILMHLRLDFMSFLE 1557
QY 228 ---VTSLD---W---N-SE---236
Db 1558 PQIISIDLPREAKTPATAPRWQPOGHKHSQWALGDVLGVNDSSQTFQDKLGSIDL 1617
QY 237 ---G---T-L---LA-T---G---243
Db 1618 QANWYRIQGGQVIDIDCLSMDTALITFTENIGNPRYAETWDFHVSSESSFLOWMMGCKP 1677
QY 244 S---Y---D---G---FA-RI---251
Db 1678 SGAHGIOLOYSLNGKDWQVLTEECVPPTIGCVHYTESSTYTTSERFQNRRTVYLPLAT 1737
QY 252 ---W---T---253
Db 1738 NSPRTFRWIOQNTYTVGADSWAIDNVILASGCPWMCGRGICDGRVCDCRGGFGPCVP 1797
QY 254 ---KD---GNL---A---S-T-L---G---263
Db 1798 VVPLPSILKDDFNGNLHPLDLPPEVYGAERGNLNGETIKSGTCLIFKGEGLRMLISRDLDC 1857
QY 264 ---Q---HK---G---P-I-F---270
Db 1858 TMTYVQFSRLRTAKTGPERSHILLOQFSVSGVTWHLMDFFYFPQTTSILFINVPLPYG 1917
QY 271 ---A---LKW---N-KK---GN---F---280
Db 1918 AQTNATRFRL-WQYNNNGKKEIWIIDDFIIDGNLNNPVLLDFTDFGPREDNWFYFG 1976
QY 281 ---I-L---SA---G-V-D---KTTII---292
Db 1977 GNIGLYCPYSSKGAPEEDSAMVFSVNEVGEHSITTRDLSVNENTIIOFEINVGCSTDSSS 2036
QY 293 ---W---DA-H---T-G-EA---300
Db 2037 ADPVRLEFRDFGATWHLLLPLCYHSSSLVSLCSTEHPSTSYAGTTQGMRRVVHFG 2096
QY 301 K---Q-O-P---P-F-H---307
Db 2097 KLHLCGSVFRWYQGYFAGSQPVTAIDNVYIGPOCEMYCGHSGCINGTKICDPYS 2156
QY 308 ---SA---PA---311
Db 2157 GPCTKISTKNPDKDDFSGQLESDFLLMGGKPSRKCGILSSGNNLPFNEGDRLMVT 2216
QY 312 ---LD---V-D---315
Db 2217 RDLDLSHARFVQFMRLGCGKGVDPDRSQPVLLQYSLNGLSWLLQEFLLFSNSNVGRY 2276
QY 316 ---MO---S---N-NT---321
Db 2277 IALEMPLKARSGSTRLRWQPSENGHPYSPWVIDOILIGNISGNTVLEDDFSTLDSRW 2336
QY 322 ---F-ASCS-TDMC-I---331
Db 2337 LLHPGQTKMPVCGSTGDALVFIEKASTRYVVTDDIADVNEBSFLQIDFAASCSTVDSYAI 2396
QY 332 ---H---V-C---334
Db 2397 ELEYSVDLGLSWHPLVRDCLPTNVECSRYHLQIRILVSDTFNKWTRITLPLPSYTRSOATR 2456
QY 335 ---K---L---G-Q---D-R---P-IK 343
Db 2457 FRWHQAPFDKQQTWAIIDNVYIGDGLCMCSGHGRCVQSCVCEQWGGLYCDEPEPESLP 2516
QY 344 T---F---Q---G-H---TN---350

2517 TQKDNFNAPSNNWLTGSGKSLTVCGAVASGLALHFGGCSRLLVTVLNLNNAEFT 2576
351 -----E-VN-AIKW-----D-----P-----359
2577 QFYFMYCCLITPSNRNQVLEYSVNGGITWNLLMEIFDQYKPGFVNILLPPDAKEIA 2636
360 T-----G-----N-L-----L-----364
2637 TRFRWQPRHGLDQNDWADNVLISGSADQRTVMDLTFSSAPVPOHSPADAGPVGR1 2696
365 A-----S-----CS-----D-D-M-----T-L-----373
2697 AFEMFLEKTSVNNENLWFHDDCTVERFCDSPDGMVLCGSHDGRVAVTHDLPPTENWIM 2756
374 -----KI-----375
2757 QPKISVCKVPEKIAQNIHQVQSTDFGVSWSLVPCQLPADPKCSGSVSQSPVFPFTEG 2816
376 W-----SM-----K-----Q-----DN-----CV-----H-D-L-QQ-----389
2817 WKRTYPLPESLTGNPVRFRFYQKYSDVQWADINFLYGLPGCLDNCGHGDCLEQICIDP 2876
390 -----H-N-K-----EI-----Y-----T-----396
2877 GYSGPNLYTHSLKTFKRFDEEIKPDLWMSLEGSTCTCGVLAENTALYFGGSTVR 2936
397 -I-----K-----W-----S-P-----T-G-----403
2937 QAITQDLRLGAKFLQYWGRIENWNTSCHRVCRKGBVLDFDGTGGLTWILLHEMDF 2996
404 -----P-G-TNN-----P-----N-----A-NLML--A-----416
2997 QKYSVRHDYILLPEGALTWTLRWQPPVINGLVSGVRAQWALDNLIGGAENP 3056
417 SA--SFD--ST-----VR-----LM-----D-----V-DR-----431
3057 SOLVDTFDDGSSHEENWSFPYNAVRTAGFCGNPSFHLWPNKCKDKTHNLSRELLIQ 3116
432 -G-----I-----C-IHT-----LTK-----H-----441
3117 PGYMMQFKIVGGEATSCGDLHSLVMELEYTKDARSQNLVQTOCLPSSNSIGCSPPQFH 3176
442 QE-PVY-----S-----VAFSPD-----G-----R-----YL-A-----458
3177 -EATYNAVNSSWKRTIQL-FDHVSSSATQFRWIKQGEETEQSWAIDHVVIGEACPK 3234
459 --SG-----SF-DKC-V--H-----I-----WNT-Q--V-----473
3235 LCSGHGYCTTGAVICIDESFGDDCSVFSHELPSYIKDNFESARVTEANWETIQGGVIGS 3294
474 -C-L--H-----YLANG-QV-----L-LN-----L-G--R--SIC--LYTL 495
3295 GCGQLAPYAHGDSLYFGCGQIROATKPLDLTRASKIMFVLQIGSPAQTDG-CNSDL-SG 3352
496 PH-----HLV-----V-----IPL-----V-----505
3353 PHTVDAKVLQYVNNNGITWHTVIAHQPKDFTQARQVSYNVPLEARNKGVLLRWQPRHN 3412
506 -----AL--IE-LLV-----LK 514
3413 GTGHQWALDHVEVLVSTRKQNYM 3437

RESULT 2

US-08-469-486-52
; Sequence 52, Application US/08469486
; Patent No. 5739281
; GENERAL INFORMATION:
; APPLICANT: Thøgersen, Hans Christian
; APPLICANT: Holtet, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: Improved method for the refolding of
; TITLE OF INVENTION: proteins

NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,486
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/192,060
FILING DATE: February 4, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06363/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 542 5070
TELEFAX: 617 542 8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 4544 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-486-52

Query Match 70.2%; Score 2562.6; DB 1; Length 4544;

Best Local Similarity 10.4%; Pred. No 3.9e-32;

Matches 414; Conservative 79; Mismatches 16; Indels 3474; Gaps 366;

QY 1 MS-----I-----SSDE-----V-----N--F-L-----11
DB 88 MSRLNGVQDCMDGSDGPHCRQLQNCRLGQCHCVPLDGTGTCVCSNPFOLQADGKT 147
QY 12 -----VY-----R-----Y-LQ-----E-----18
DB 148 CKDFDECSVYGTCSQLCTNTDGSFICGCVGYLLOPDNRSCKAKNEPVDPPVLLIANQ 207
QY 19 -----SG-----FS-----H--SA-----F 26
DB 208 NILATVLSAQVSTTPTSTRTQTAMDFSVANETVCWVHVGDSAAQTOLKCARMPGLKGF 267
QY 27 -----TEGIK-S-HISQ-----S-N-----I-----NG-----41
DB 268 VDEHTNISLSLHVHQAIDMLTGNFYFVDDIDDRIFVNCNRNGDTCVTLDDLELYNPKG 327
QY 42 -AL-----VP-----P-----A-A-48
DB 328 IALDPAMGVFFTDYQGIKPKVERCMDGQNRKLVDSKIVFPHGITLDLVSLVYWADAY 387
QY 49 --LJ-----SLIQG-L-Q-Y--V-E-----A-EVS-I-----N 67
DB 388 LDYIEVDYVEGKGRQTIIO-GILIEHLYGLTFVENLYATNSNANAQKTSIVNRFN 446
QY 68 --E-----D-GTL-----R-----72
DB 447 STEYQVTVTRVDKGGALHIYHORQPRVSHACENDQYKPGGSDICLLANSHKARTCRC 506
QY 73 ---F---DG-----R-----PIE-----SL---81
DB 507 RSGFSLGSDGSKCKPEHELFLVYKGRPGIIRGMDMGAKVDPDEHMIPIENLMNPRALDF 566

QY 82 -----S-LI-----D-----AV-M-----PD-- 90
Db 567 HAETGYFADTTSYLGRQKIDGTERETILKOGIHNHVEGVAVDMGDNLYWTDDGPKKT 626
QY 91 V-V-----QTR-----QO----- 97
Db 627 ISVARLEKAAQTRKTLIEGKMTWHPRAIVDPPLNGMWTWDEEDPKDSRRGRLERAMWDG 686
QY 98 -----A-Y-R-----D-K----- 102
Db 687 SHRDIFVTSKVLWPNGLSLDIPAGRLYWDAFYDRIETILLNGTDRKIVVEGPELNHAF 746
QY 103 -----L-----AQOAA----- 109
Db 747 GLCHGNYLFWTEYRSGSVYRLRGVGGAPPTVTLRSPPIPEIRMYDAQOQOVGTNK 806
QY 110 -----AA-----A-----A-AASQ- 120
Db 807 CRVNNGGSSILCLATPGSRQCAEAQVLDADGVTCLANPSYVPPQCPQGEFACANSRC 866
QY 121 Q-----G-----S-----A-----K-----N-G-----E 128
Db 867 IQERWKCDGNDCLDNDSEAPALCHQHTCPDRFKCENNRCPNRLWLCGDNDGNDGSEDE 926
QY 129 N-T-A-----N-----G-----E-----E-----N 136
Db 927 SNATCSARTCPPNQFSCASGRCPISWTCBLDDDCGRSDSASACAYPTCPPLTQFTCNN 986
QY 137 G-----A-H-----T-----I----- 141
Db 987 GRCININWCDNDGNDGNDSEAGCSHSCSTQPKNSGRCPPEHWTCDGNDGNDGYSDE 1046
QY 142 --AN-N-----HTD-----M-----M-----E-----V-D- 152
Db 1047 THANCTNQARTPPGGCHTDFQCRLDGLCPLRWCRDGTDCMDSDSEKSCGVTHVCDP 1106
QY 153 --G-D-----V-----E-I-----PP-----NKAV-----V- 164
Db 1107 SVKFGCKDSARCISKAWCGDNDCEDNDSBENCESLACPPSHPCANNNTSVCLPDPKLC 1166
QY 165 -----L-R-----G-H-----E-----S-----E-----V----- 172
Db 1167 DGNDDCGDSGDEGLCDQCQSLNNGCCHNSVAPGEGIVCSPLGMLGPDNHTCQIQSY 1226
QY 173 -----F-I-CA-----W-----N----- 178
Db 1227 CAKHLKCSQKCDQNKFSVKSCYEGWVLEPDGSCSLDPKFPFIIFSNRHRIRIDLHK 1286
QY 179 -----P-----V-----S-----D-----LL-----A-----S----- 186
Db 1287 GDYSVLVPLGRLNTALDFHLSQSALYWTVDVVEDKIYRGKLLDNGALTSFEWVIQYGLATP 1346
QY 187 -----G-----S-----GD-----S-----T----- 192
Db 1347 EGLAVDWIAGNIYVESNLQIEVAKLDGTLRTLLAGDIEHPRAIALDPRDGLFWTDW 1406
QY 193 A-RI-----W-N----- 197
Db 1407 DASLPRIEAAASMGAGRRTVHRETGSGGWPNGLTVDYLEKRILMIDARSARYDGS 1466
QY 198 -----LS-----E-----NST-----SG-----S-TQ-----L 209
Db 1467 GHMEVLRGHFLSHPPFAVTLYGGEVYWTDRWTN-TLAKANKWTGHNVTVVQRTNQPFDL 1525
QY 210 V-----L-----R-----H-----C----- 214
Db 1526 QVYHPSRQPMAPNCPCEANGGQPCSHLCLINYNRTVSCAPHLMLKHKDNTTCVEPKFL 1585
QY 215 -----I-----RE----- 217
Db 1586 LYAROMBEIRGVDLAPYNYIISFTVPDIDNVTLVDYDAREQRYVNSDVRTQAIKRAFIN 1645

QY 218 G-G-----ODVPSNK-----D-V-----TS-----LD----- 232
Db 1646 GTGVETVVSADLP-NAHGLAVDWVNRNLFWTSTDTNKKQINVARLDGSKFNNAVQGLEQP 1704
QY 233 -----W-----N-S-----EG-----TLL-----A-----TG-----S-Y----- 245
Db 1705 HGLVHPLRKLKWTGDNDISMANMGDSNRTLLFSQKQGPVGLAIDFPESKLYWISSGNH 1764
QY 246 -----D-GF-----A-R-----I-----W-----T-----K-DG----- 256
Db 1765 TINRCLDGSGLVEIDAMRSQKATATAIMGDKLWADQVSEKMGTCRADGSGSVLR 1824
QY 257 N--LA-----S-----T-----L-----GQ----- 264
Db 1825 NSTLWVHMKYVDESITQDHKGTNCPNVNNGDCSQLCLPTSETTRSCMCTAGISLRSQQ 1884
QY 265 -----HKG-----P-----I-P-A-----L----- 272
Db 1885 ACEGVGSFLLYSVHEGIRGIPLOPNDKSDALVPVSGTSLAVGIDFHAENDTIYVMDGLS 1944
QY 273 -----K-----W-----N-----K-----KGN-----P-I-----L-----S 283
Db 1945 TISRAKRDQWREDVVNTGIGRVEGIAVDWIAWNIYWTQDQFVIEVARLNGSFRYVIS 2004
QY 284 AGVDK-----T-----T-----I-----I----- 292
Db 2005 QGLDKPRAITVHEKGYLFWTEWQYPIERSRLDGTERTVVLVNSISWNGISVDYQDG 2064
QY 293 --W-DA-----H----- 296
Db 2065 KLYWCARDTKIBRIDLETGENREVLSNNMDFSVSFEDFIYWSDRTHANGSIKRG 2124
QY 297 -----TG-----E-----AK-----OQ----- 303
Db 2125 KDNATSDVPLRTGIGVQLKDIKFVNRDRQGTNVCVAVANGCQOCLLYRGRGORACAH 2184
QY 304 -----P----- 304
Db 2185 GMLAEDGASCREVAGVLLYSERTILKSIHLSDERNLAPVOPFEDPEHMKVIALAFDYR 2244
QY 305 -----P-----F-----H-----S-----A----- 309
Db 2245 AGTSPGTNPNRIPFSDIHFQNIQIINDGSRITIVENVSGVEGLAYHRGMDTLWTSYTT 2304
QY 310 -----P-AL-----D-----V-D-----W-O-----S 318
Db 2305 STITRHTVQTRPGAPERETVITMSGDDHPRAEVLDECONLMFWTNNEQHPSIMRAALS 2364
QY 319 --N-----N-----T-----F-AS-----C-----S----- 326
Db 2365 GANVLTLEKIDRTPNGLAIDHRAEKLYFSDATLDKIERCEYDGSYHYVILKSEPHVFPF 2424
QY 327 -----TD-----M-----C-I 331
Db 2425 LAVYGEHFWTDVRRRAVQRANKHGVSNMKNLLRVDIPQOPMGI IAVANDTNSCELSPCI 2484
QY 332 -----HV-----CKLG-----QD-----R----- 340
Db 2485 NNGGQDCLLTHQGHVNCSCR-GGRILQDLDLTCRAVNSSCRAQDEFEANGECINFSLT 2543
QY 341 -----P-----I-----KT 344
Db 2544 CDGVPCHKDKSDEKPSYCNRRCKTFRQCSNGRCVSNMLWMCNAGDDCGDSEI PCNKT 2603
QY 345 -----FQ----- 346
Db 2604 ACGVEFRCDGTGICIGNSRQNFVDCEDASDEMNGSATDCSSYFRLGVKGLFQPCERT 2663
QY 347 -----G-----H 348
Db 2664 SLCYAPSWVDGANDCGDYSYDERDCFGVKRPRCPPLNYFACPSGRCPIMSWTCKEDDCEH 2723
QY 349 --T-N-----EV-----N-A-I-K-W-----D-----PT-----G 361

Db 2724 GEDETHNCFCSAQFEQNHRCISQWLCGSDGDDCGSDEAAHCEGKTCGSPSFCFG 2783
Qy 362 N-----L-----LAS-C--S--DD--M-----371
Db 2784 THVCVERMLCDGDKCAGADESIAAGCLYNSTCDDREBFMCNQRCIPKHFVCDHRDC 2843
Qy 372 -----T-----Lk-----I--W-----S--MK 379
Db 2844 ADGSDSPCEYPTCGSPSEFCANGRLSRQWECDECHDQSDAPKPNHCTSPK 2903
Qy 380 -----QD-----381
Db 2904 CNASSQFLCSSGRVAEALLCQDGDGSSDERGCHINECLSRKLSGCSQDCEDLKIGF 2963
Qy 382 -----N-----CV-----H-----385
Db 2964 KCRPGFRLKDGRTCADVDECSFTFPCSORCINTHGSYKCLCVBGPYAPRGDPSCKA 3023
Qy 386 -----D-----L-Q-----Q-----389
Db 3024 VTDEEPFLFANRYLRKLNLDGNTLLKQGLNNAVALDFDYREQMIYWDVTTOGSMI 3083
Qy 390 -----H-----N-----K-----EI-----Y-TI-----397
Db 3084 RRMHLGNSVQVLRGLSNDPLGLAVDWGGLNYWCDKGRDTIEVSKLNGAYRTVLVSSG 3143
Qy 398 --K-----W-----S-----P-----T-----402
Db 3144 LREPRALVVDVQNGYLYWTDGPHSLIGRIGMDGSSRSVIVDTKITPNGLTLDYVTERI 3203
Qy 403 -----G-----P-----T-K--H-----Q-----442
Db 3204 YWADAREDYEFASLDGSRHVVLSQDIPHI FALTIFEDYVYWDNETKSNINRAHKTGT 3263
Qy 407 N-----N-P--N--ANL-MLAS-----A--S-F- 420
Db 3264 NKTLTLLTLRPMDLHVHALRQDPVNPHEPKVNGGCSNLCCL-SPGGGHKCACTNFY 3322
Qy 421 --D-----S--T--V--R--LW--DV-----D-----R-----431
Db 3323 LGSGRGTCVSNCTASQFVCKNDKCIPIFWMKCDTDDCGDSHSDPDPCEPKRPGQFQCS 3382
Qy 432 -GIC-----IHT-L-----T-K--H-----Q-----442
Db 3383 TGICTAPFICDGNDCQNSDEANDCIHVCLPSQFKCTNTRNCIPIFRGNCQDNCGDG 3442
Qy 443 -----E-----P-VY-----S-----V-AF-- 450
Db 3443 EDERDCPEVTCAPNQFQCSITKRCIPRVWVCDRNDNCVDGSDPANCTQTCGVDFRCK 3502
Qy 451 -S-----P-----DG-----R--Y-----LAS--G 460
Db 3503 DSGRCIPARWKCDGEDCGSDGSDPEKCECDERTCEPYQFRCKNNRVCPGRWQCDYDNDG 3562
Qy 461 -----S-F-----D-K-CV-----466
Db 3563 DNSDEBSCTPRPCESEFSCANGRCIAGRWKGDHDCDGSDEKDCPTPCMDMDQFQCKS 3622
Qy 467 -H-I--W-----NT-----471
Db 3623 GHCIPLRWRCADADCMWGSDEACGTGVRTCPDLDFQCNTTLCKPLAWKCDGEDDCGN 3682
Qy 472 -----QVCL-----H-- 476
Db 3683 SDENPECARFVCPNRPFRCKNDKRCIWTGRQCDGTDNCGDGTDEBCEPPTAHTHCK 3742
Qy 477 -----YL-----N-----479
Db 3743 DKKEFLCRNQRCLSSLRNCMFDGCGSDEEDCSIDPKLTSCATNASIGDBARCVTE 3802
Qy 480 -----GQ-----V--L-----L--NL--G-----R--S--IC-- 491

Db 3803 KAAYCACRSGFHTVPGQGCQDINECLRFGTCSQLCNNTKGGHLCSCARNFMKTHNTCKA 3862
Qy 492 -----LV-----493
Db 3863 EGSEYQVLIADDDNEIRSLFPQPHPSAYEQAFQCGDESVIDAMDVHVHAKGRVYVWTHWHTG 3922
Qy 494 T-----LP-----H-----HL-----V-----VI 502
Db 3923 TISYRSLPAAPTTSNRHRRQIDRGVTHLNISGLKMPRGIAIDWVAGNVYWTDSGRDVI 3982
Qy 503 -----P-LVA--L-----IE-----L--LV-- 512
Db 3983 EVAQMKGENRKTLSIGMIDEPHAIIVDPLRGTMWSDWGNHPKXIETAAMDGTLETLLVQD 4042
Qy 513 -----L-----K 514
Db 4043 NIQWPTGLAVDYHNERLYWADAK 4065

RESULT 3
US-08-469-658-52
; Sequence 52, Application US/08469658
; Patent No. 5917018
; GENERAL INFORMATION:
; APPLICANT: Th egersen, Hans Christian
; APPLICANT: Holtet, Thor Las
; APPLICANT: Eterodt, Michael
; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,658
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/192,060
; FILING DATE: February 4, 1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06363/002002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 542 5070
; TELEFAX: 617 542 8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4544 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-469-658-52

Query Match 70.2%; Score 2562.6; DB 2; Length 4544;
Best Local Similarity 10.4%; Pred. No. 3.9e-32;
Matches 414; Conservative 79; Mismatches 16; Indels 3474; Gaps 366;
Qy 1 MS-----I-----SSDE-----V-----N--F-L-----11

```
Db 88 MSRLCNVQDMDGSDRGPHCRELOQNCRLGCOHCHVPTLDGPTCYCNSSFQLOADGKT 147
Qy 12 -----VY-----R-----Y-LQ-----E-----18
Db 148 KDPDECSVYGTCSQLCTNTDGSFICGCVGYLLQPDNRSCAKNEPVPDPVLLIANSQ 207
Qy 19 -----SG-----PS-----H-SA-----F 26
Db 208 NILATYLSGAQVSTITPTSTROTMTAMDFSYANETVCWVHVGDSAAQTOLKACRMPGLKGF 267
Qy 27 -----TFGIK-S-HISQ-----S-N-----I-----NG-----41
Db 268 VDEHTINISLSLHVEQMAIDLWLTGNFYFVDDIDRRIFVCNRNGDTCVTLLDLELYNPKG 327
Qy 42 -AL-----VP-----P-----A-A-48
Db 328 IALDPAMGVFTDYQGIKPKVERCDMDGQNRKTLVDSKIVFPHGITLDDLVSRLVYWADAY 387
Qy 49 --LI-----STIQK-L-Q-Y--V-E-----A-EVS-I--N 67
Db 388 LDIEVVDYEGKGRQTIQ-GILIEHLYGLTVFENYLYATNSDANAAQKTSIVRNRFN 446
Qy 68 -E-----D--GTL-----72
Db 447 STEVQVTVRVDKGGALHIYHQRQPRVRSACENDQYKPGGCSIDICLLANSHKARTCRC 506
Qy 73 -F-----DG-----R-----PIE-----SL--81
Db 507 RSGFSLGSDGSKCKPBHEFLVYKGRGPIIRGMDMGAKVDPDEHMIPIENLMPALDF 566
Qy 82 -----S-LI-----D-----AV-M-----PD--90
Db 567 HAETGYFPADTYSYLGROKIDGTERETILKOGIHNVGVAVDMGDNLVWTDGPKKT 626
Qy 91 V-V-----QTR-----QO-----97
Db 627 ISVARLEKAAQTRKTLLEGKMTHPRAIVDPVPLNGWYMTDWEEDPKDSRRGRLERAWMDG 686
Qy 98 -----A-Y-R-----D-K-----102
Db 687 SHRDIFVTSKVLWPNGLSLDIPAGRLYWDAFYDRIETILNGTRDKIVYEGPELNHAF 746
Qy 103 -----L-----AQOAA-----109
Db 747 GLCHGNYLFWTEYRSGSVRLRGVGGAPPTVLLRSERPPIFEIRMYDAQOQVGTNK 806
Qy 110 -----AA-----A-A-----A-RAAQ-120
Db 807 CRYNNGCSSILCLATPGSRQCAEADQVLDADGVTCLANPSYVPPQCPQGFACANSRC 866
Qy 121 Q-----G-----S-----K-----N-G--E 128
Db 867 IOERWKGDGNDCLDNDEAPALCHQHTCPSDRPKCENNRCPNRLWLCGDGNDGNSDE 926
Qy 129 -N-T-A-----N-----G-E-----E--N 136
Db 927 SNATCSARTCPNPQFSCASGRCTPISWTCBLDDCGDRSDESASCAYPCTCPPTQFTCNN 986
Qy 137 G-----A-H-----T-----I-----141
Db 987 GRCININWRCDNDGNDGNSDEAGCSHCSSTQPKCNSGRCPBEHMTCDGNDGNDGYSDE 1046
Qy 142 --AN-N-----HTD-----M-----M-E--V-D-152
Db 1047 THANCTNQATPPGGCTHDFQCRLDGLCPLRWCRDGDTCMDSDSKSCGCVTHVCDP 1106
Qy 153 -G-D-----V-----E-I--PP-----NKAV-----V-164
Db 1107 SVKFGCKDSARCTSKAWVCGDNDCEDNDSBENCESIACRPPHPCANNTSVCLPPDKLC 1166
Qy 165 -----L-R-----G-H-----E-S-----V--172
```

QY 310 P-AL D V-D W-Q S 318
Db 2305 STITRHTVQTRGAFERETVITMSGDHPAPVLDCEQNLMEFTWNNQHPHSIMRAALS 2364
QY 319 N N T P-AS C S 326
Db 2365 GANVLTLEIKDITPGLAIDHRAEKLKLYFSDATLDKIERCEYDGSRYVILKSEPVHPFG 2424
QY 327 TD N M C-I 331
Db 2425 LAVYGEHIFWTDWVRVAVRANKHVSNMKLLRVDIPQPPMGLIAVANDTNSCELSPCRI 2484
QY 332 HV CKLG QD R 340
Db 2485 NNGGCQDLCLTHQGHVNCSC-GRILQDLTCAVNSCRAQDEFECANGBCINFSLT 2543
QY 341 P I I--KT 344
Db 2544 CDGVPCHKDSEKPSYCNRRCKTFRQCSNGRCVSNMLWCNAGDCCGDSDEIPCNKT 2603
QY 345 G V F Q 346
Db 2604 ACVGFRCDGTICGNSRNCQFVDCEDASDEMNCSATDCSYFRLGVKVLFPQCERT 2663
QY 347 G H 348
Db 2664 SLCYAPSWCDGANDGCDYSDERDCPGVKRPRCLNYFACPSGRCLPMSWTCKEDDCEH 2723
QY 349 T N EV N-A-I-K-W D PT G 361
Db 2724 GEDETHCNKFCBAQECQNHRCISQWLCDGSDGDEAAHCEGKTCGSPSPFCG 2783
QY 362 N LAS-C S-DD M QD 371
Db 2784 THVCVPERLWCDKDCADGADBSIAAGCLYNSTCDDREFMCONROCIPIKHVFCDHRC 2843
QY 372 T LK I W S-MK 379
Db 2844 ADGSDSPCEYPTCGSPFRCANGCLSRQWECDENDCHDSDEAPKNPHCTSPCHK 2903
QY 380 C NASSQFLCSGRVAEALLCNGQDCGSDGSDRGCHINECLSRKLSGCSQDCDLKIGF 2963
Db 382 H CV H 385
Db 2964 KCRPGRFKDGRTCADVDECSFTTFCQRCINTHGSYKCLVGEYAPRGDPSCKA 3023
QY 386 D L Q Q 389
Db 3024 VTDEEPFLIFANRYLRKLNLDGNTLLKQGLNNAVALDPDYREQMIYTDVTTQSGMI 3083
QY 390 H N K EI Y-TI 397
Db 3084 RRMHLNGSNVQLHRTGLSPDGLAVDWGNGLYWCDKGRDITIEVSKLNGAYRTLVSSG 3143
QY 398 K W S P T 402
Db 3144 LREPRALVVDVQNGYLYWTDWGDHSLIGRIGMDSSRSVIVDTKIWPNGLTLDVYTERI 3203
QY 403 G P P 406
Db 3204 YWADAREYIEFASLDGSRHVVLSQDIPHI FALTIFEDYVYVWTDWETKSNRAHKTGT 3263
QY 407 N P N ANL-MLAS A-S-F 420
Db 3264 NKTLLISTLHRPMDLHFHALRPDPVNPKNVNGCSNLCLL-SPGGGHKACACPTNY 3322
QY 421 D S-T V R-LW DV R 431
Db 3323 LGSDBGRTCVSNCTASQVCKNDKCFPFWKCDTDDCGDHSDEPDCBPKCRPGQFQCS 3382

QY 432 -GIC IHT-L T-K-H Q 442
Db 3383 TGICTNPATFCGDNDQDNDSEANCDIHVCLPSPQFKCTNTRNCIPFRNCQDNGDG 3442
QY 443 E P-VY S V-AP 450
Db 3443 EDERDCPEVTCAPNOFQCSITKRCIPRVVWDRDNDVDSDEPANCTQMTQGVDFRCK 3502
QY 451 S P DG R Y LAS-G 460
Db 3503 DSGRCIPARWCKDGEDDCGSDGDEPKKECDERTCEPYQFRCKNNRVCVPGRWQCDYDNDG 3562
QY 461 S F D-K-CV 466
Db 3563 DNSDEBSCPTPCSESEFSCANGRCIAGRWCQDGDHCDGSDKDCOTPRCDMDQFOCKS 3622
QY 467 H-I W NT 471
Db 3623 GHCIPLRWCADADCMGSDDEACGTGVRTCPLDFEQCNNTLCKPLAWKCDGEDDCGN 3682
QY 472 QVCL H 476
Db 3683 SDENPEECARFVCPNRPFRCKNDRVCLWIGRQCDGTNCGDGTDEEDCEPPTAHTHCK 3742
QY 477 YL N 479
Db 3743 DKKEFLCRNQRCLSSLRNMFDDCGSDGDEEDCSIDPKLTSCATNASICGDEARCVTE 3802
QY 480 QO V L NL-G R S-IC 491
Db 3803 KAAACACRSQFTVPGQCODINECLRFGTCSQLCNNTKGHLCSARNFMKTHNTCKA 3862
QY 492 LY L 493
Db 3863 EGSEYQVLIADNEIRSLFPCHPHSAYEQAFQGDSESVRIDAMDVHVKAGRVYTWHTG 3922
QY 494 T LP H HL V VI 502
Db 3923 TISYRSLPAPPTTSNRHRRQIDRGVTHLNI SGLKMPRGIAIDWAGNVYWTDSGRDVI 3982
QY 503 P-LVA-L IE L-LV 512
Db 3983 EVAQMKGENRKLISGMIDEPHAI VVDPLRGTWYSDWGNHPKPIETAMDGLRLTLVQD 4042
QY 513 K 514
Db 4043 NIQWPTGLAVDYHNERLYWADAK 4065

RESULT 4

US-08-476-515A-84

; Sequence 84, Application US/08476515A

; Patent No. 6239270

; GENERAL INFORMATION:

; APPLICANT: Akerstrom, Goran

; APPLICANT: Juhlin, Claes

; APPLICANT: Raak, Lars

; APPLICANT: Crumley, Gregg R.

; APPLICANT: Morse, Clarence C.

; APPLICANT: Murray, Edward M.

; APPLICANT: Hjalms, Goran

; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments

; NUMBER OF SEQUENCES: 84

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Martin Savitzky

; STREET: Rhone-Poulenc Rorer Inc., 500 Arcola Rd.;

; CITY: Collegeville

; STATE: PA

; COUNTRY: USA

; ZIP: 19426-0107

; COMPUTER READABLE FORM:

Db 1620 DYMFCDYNGHRRQVIASDLIRHPYALTFLFEDSVYVWTDTRATRRVMRANKWGHGNSVV 1679
Qy 220 -----Q----- 220
Db 1680 MYNIQWPLGIVAVHPKQNSVNPFCASRCSHLCLSSQGFPHFYSCVPCSWLSPLLN 1739
Qy 221 -----V-----P--SN-K-----DV----- 228
Db 1740 CLRDDQFLITVRQHIIIFGISLNEVKSNDAMVPIAGIQLGDLVEFDADAEQIYVWENPG 1799
Qy 229 -----T-----S-----LDW-----N-----S-E-G-----T 238
Db 1800 EIHVKVTDGNTVRVFAISVMGSPSMLALDWISRNLYSTNPTQTSIEVLTHGDIRVKT 1859
Qy 239 LLA--T-----G-S-----Y-----DG-----F----- 248
Db 1860 LIANDGTALGVGPIGITVDPARGKLYWSDQDTSVPAKIASANMDGTSVKTLFTCNLE 1919
Qy 249 -----A-----R-I-----W-----T----- 253
Db 1920 HLECVTLIDIEQKLYAVTGRGVIERNVGDTRMILVHQLSHPWGIAVHDSFLYTTDEQ 1979
Qy 254 ---K-D---G-N---L---ASTL-G-Q-H-----K----- 266
Db 1980 YEVIERVDKATGANKIYLNDVNPRLGLQVYHRRNAEASSGCSNNNACQOICLPVPGG 2039
Qy 267 -----G-----P-----I--PALK----- 273
Db 2040 LFSCACATGFLNPDNRSCFPYNSFIVVMSLSAIRGFSLSHDSHSETWVPVAGGRNALH 2099
Qy 274 -----W-----N-----K-----GN----- 279
Db 2100 VDVDVSSGFYWCDFSSVASDANAIIRIKPDGSSLMNIVTHGICENGVRGIAVDWVAGNL 2159
Qy 280 ----- 280
Db 2160 YFTNAFVSETLIEVLINTTYRRVLLKVTDMPRHIVVDPKNRYLFWADYQRPKIERSF 2219
Qy 281 -----I-----L-----S-A-----G----- 285
Db 2220 LDCNRTVLVSEGI VTPRGIAVDRSDGYVWVDSLDIIARIINGENSEVIRYGSRYPT 2279
Qy 286 -----VD-----K-----T---TII-----W-----D----- 294
Db 2280 PYGITVPFENSIIWDRNLKXIFQASKEPENTPEPTVIRDNIINLRDVTIFDKVQVRSPA 2339
Qy 295 -----A-----HT-----G-----E-----A----- 300
Db 2340 EVNNPCLENNCGCSHLCPALGHLTPKCDCAFGLQSDGKNCAISTENFLIPALSNLSR 2399
Qy 301 -----K-----Q-----O-F-----P----- 305
Db 2400 SLHLDPENHSPPTQTNINVERTVMSLDVSDRIYFTQNLASGVQISYATLSGHIHTPT 2459
Qy 306 -----F----- 306
Db 2460 VIASGIGTAGIAFDWITRIYSDYLNQMINSMAGDSNRTVIARVPKRAIVLDPCCQ 2519
Qy 307 -----H-----S--AP--ALD----- 313
Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSSLVMPSGLTLDYEEDLLYVWDASLQRIERST 2579
Qy 314 ---VD---W---Q---S-----N----- 319
Db 2580 LTGVDRVIVNAVAHAFGLTYGYIYVWTDLYTQRIYRANKYDGSQGIAMTTLNLSQPRG 2639
Qy 320 -NT-----F----- 322
Db 2640 INTWVKNQKQCNPCQFNGGCSHICAPGNGAEQCPHEGNYWLANNEKHCIVDNGER 2699
Qy 323 --AS---CS-----TD-M---C-IH-----V----- 333
Db 2700 CGASSFTCSNGRCISBEWKCDNDNDGSDGSESVICALHTCSPTAFTCANGRCVQYSYR 2759

Qy 334 CKL-----G--QD-----R--P--I----- 342
Db 2760 CDYNDGDSDEAGCLFRDCNATTFMCMNRRCIPREFICNGVDNCHDNTSDKNCPSD 2819
Qy 343 -----K-----T-----FQ----- 346
Db 2820 RTCQSGYTKHNSNICIPRYVLCDGNDGNSDENPTYCTTHTCSSSEFQCASGRICIPQ 2879
Qy 347 -----GH--T----- 349
Db 2880 HWYCDQETCFDASDPASCCHSERTCLADEFKDGRGRCIPSEWICDGDNDGMSDEDK 2939
Qy 350 -----N---E---VNA-----IK--W----- 357
Db 2940 RHQCQNQNSDSFELCVNDRPPDRRCIPQSWVCDGVDCDGDYDENQNCRTTTCSENEPT 2999
Qy 358 ----- 357
Db 3000 CGYGLCIPKIFRCDRHNDGDSYDERGCLYQTCQNOFTCQNGRCISKTFVCDENDDCGD 3059
Qy 358 -D-----PT-----G-----N-----L----- 363
Db 3060 GSDELMHLCHTPEPTCPPHSFKCDNGRCITEMMKLCNHLDDCLDNDSEKGGCINECHDPSI 3119
Qy 364 -----LAS--CS-----D-D---M----- 371
Db 3120 SGCDHNCITLTSFYCSRGYKLMDSKRTCDVIDECTEMPFVCSQKCVENGVSYICKCA 3179
Qy 372 ----- 371
Db 3180 PGLREPDGKTQRONSIEPYLIFSNRYLRLNLTIDGYFYSLILEGLDNVVALDFDRVEK 3239
Qy 372 -----T-----L---K---I-----W-----S--- 377
Db 3240 RLYWIDTORQVIERMFLNKTNETIINHRLPAEASLAVDMVSRKLYWLARDLGLFVSXL 3299
Qy 378 -----M-KQ---D-N--C-----VHDLQ-----H-----NK- 392
Db 3300 NGHRRMLAQHCVDANNTFCFDPNPRGLAH--POGYLYWADMGHRAIYGRVGMGTNKS 3357
Qy 393 -----E-----I-YT-----IK----- 398
Db 3358 VIISTKLEWPNGITIDYTNDLLYWAHAHLGYIEYSDLEGHHRHTVYDGLPHFPATIFE 3417
Qy 399 -----W-----S-P--T---G--- 403
Db 3418 DTIYWTDMNTRTVEKNGKYGDSNRQTLVNTTHRPFDIHVYHPYRQPIVSNPCGTNNGCS 3477
Qy 404 -----PG-----T-----NN----- 408
Db 3478 HLCLIKPGKGFTCECPDDFRTLQLSGTYCMPMCSSQFLCANNKRCIPHWKCDGQKD 3537
Qy 409 -----P-----NA--N-----LML-----AS 417
Db 3538 CSDGSDALALCPORFCLGQFCQDGNCTSPOTLCNAHQNCPCDGSDEDL--LCENHHDS 3596
Qy 418 ---A---S---F---DS---T---VR---L---W---DV 429
Db 3597 NEWQCANKRCIPESWQCDTFNDCEDNDSBDSHSCASRTCRPGQFRANGRCIPQAWKCDV 3656
Qy 430 ---DRG---I--CI--H-----T-L----- 438
Db 3657 DNDGDSDEPTEECSSAHLCDNFTEFSCKTNRYCIPKAVCNVGVDDCRDNDSEQCEE 3716
Qy 439 -T-----K-H-----Q-----E---P-----V----- 445
Db 3717 RTCHPVGDFRCNMHHCIPLRWQCDQNDGNDSENCAPRECTESEFRVCNQCIPSRW 3776
Qy 446 -----Y---S---V-----A---F 450
Db 3777 ICDHYNDGDSNDRDCMERTCHPEYFOCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836

QY 451 SPGRYL-AS-----GS-----F----- 462
Db 3837 -PDGAYCATMFECKNHVICPPYWKCGDDDCGSDGDEHLCLDVPKSPNFRCDNR 3895
QY 463 -----D-----K-----CV-H-I----- 468
Db 3896 CIYSHEVCNGVDCGDTDEBHCRTKPTKPTCEYKCGNGHCIPHDNVCDADDCGD 3955
QY 469 W-----N-TQ-----V-----CL----- 475
Db 3956 WSDLGCKGKERTCAENICEQNTQNEGFCCTAGTETNVFRTSCLDINECEQFG 4015
QY 476 -----H----- 476
Db 4016 TCPOHCRNTKGSVEVCADGFTSMDSRPGKRCABGSSPLLLPDNVRIRKYNLSSEFS 4075
QY 477 -YL-----N-----G----- 480
Db 4076 EYLQDEEYIQANDYDWDKIDGLSVVYTYVRGBSGRGAIKRAVIPNFESGRNVLQVEVD 4135
QY 481 -----Q-----V-----L-----N----- 485
Db 4136 LKLKYNMOPDGLAVDWGRHYWSDVKNKRIEVAKLDRYKWLITDLDQPAIAVNPX 4195
QY 486 LG-----R-----SI----- 490
Db 4196 LGLMFTWCKEPIESAWNGEDRNILVFDLGNWPTGLSIDYNDRIYWSDFKEDVIET 4255
QY 491 ----- 490
Db 4256 IKYDGTDRRTAKEAMPYSILDFEDQLYWISKEGVEWKNKFGQKKEKTLVNPWLT 4315
QY 491 -----CL-----Y-----T----- 494
Db 4316 QVRIHQRYNKSVPNLCKQICSHLCLLRPGYSCACPOGSSFTIEGSTECDAAIELPIN 4375
QY 495 LP-----H----- 497
Db 4376 LPPRCRCHGNCYFDETDLPKCKPGSYGKYKCEMAFSGISPGTTAVAVALLTILLIV 4435
QY 498 -----H-----LV-----V-----I-----P 503
Db 4436 IGALAIAGFYHRTGSLPALPKPLSLSLVPSSENGVTFRSGADLNMIDIGVSGFGP 4495
QY 504 -----LVA-----LI-E-LL-----VLK 514
Db 4496 ETAIDRSMASEDFVMGKQPIIFENPMYSARDSAV-K 4533

RESULT 5
US-08-652-877-84
; Sequence 84, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Rask, Lars
; APPLICANT: Crumley, Gregg R.
; APPLICANT: Morse, Clarence C.
; APPLICANT: Murray, Edward M.
; APPLICANT: Hjaln, Goran
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd., 3C43
; CITY: Collegeville
; STATE: PA USA
; COUNTRY: USA
; ZIP: 19426-0107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: Macintosh
OPERATING SYSTEM: System 7.5.1
SOFTWARE: Word 6.0 (Patentin)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/652,877
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: PCT/US95/15203
FILING DATE: 22-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,836
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/487,314
FILING DATE: 07-JUNE-1995
ATTORNEY/AGENT INFORMATION:
NAME: Savitzky, Martin
REGISTRATION NUMBER: 29,699
REFERENCE/DOCKET NUMBER: A1355E-US
TELEPHONE: 610-454-3816
TELEFAX: 610-454-3808
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 4655 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-652-877-84

Query Match 70.1%; Score 2558.8; DB 3; Length 4655;
Best Local Similarity 9.5%; Pred. No. 5.3e-32;
Matches 430; Conservative 60; Mismatches 18; Indels 4032; Gaps 371;
QY 1 M-----S-----I-----S--SDE- 7
Db 1 MDRGPAAVACTLLALVACLAPASGQECDSAHFRGSGHCPADWRCDTGCKDCDDADEI 60
QY 8 ---VN---F---L---V----- 12
Db 61 GCAVTVCOQGYFKQSEGQCIPSSWVCDQDQDCDDGSDERQDCSQSTCSSHQITCSNGQC 120
QY 13 ---YR-----Y---L---Q-----E---S 19
Db 121 IPSEYRCDHVRDCPGADENDCOYPTCEQLTCDNGACYNTSKCDKWKVDCRDSSEINCT 180
QY 20 ---G-FS-----H---S---A---F---T---F---G---I----- 30
Db 181 EICLHNEFCGNGECIPRAYVVDHNDQCGSDHACNYPTCGGYQFTCPSGRCIQNVW 240
QY 31 ---K-----S---H-----IS-----Q--SNI 39
Db 241 CDGEDCKNGDEDCGSGPHDVHKCSPREMSCPSGRCISIVKVCDDGILDCFGREDEN- 299
QY 40 N---G-----AL-----VP-----PAALI-----S-----I---I-- 53
Db 300 NTSTGYCSMTLCSALNCQYQCHETPYGGACFCPPGYIINHNDSTRCTVEFDQCIWIGICD 359
QY 54 QK-----G---L---QY-----VE 61
Db 360 QKCESRPGRLCHCEBGYILRGQYCKANDSFGESAIIFSNGRDLLIGDIHGRSFRILVE 419
QY 62 ---A-----E---VSI-----N 67
Db 420 SQNRGAVGVAFHYHLQRFVFTDTQNKVFSVDINGNLQIVLNVSVETENLAVDWNN 479
QY 68 ---E-----DG---TL----- 72
Db 480 KIYLVETKVNRIKIDWNLGDSYRVTLITENLHGRGIAVDFTVGYLFFSDWESLSGEPKLE 539
QY 73 --F-DG--R-----P-----IE-----S 80

Db 540 RAFMGDSNRKDLVKTLGNPAGVTLDMISKRVYVWDSRFDYIETVTYDGIQRKTVVHGS 599
QY 81 L-----SL-----I-----D-----AVM----- 88
Db 600 LIHPFGVSLPEGQVFTDWTKAVLKANKFETNPQVYQASLRPGVTVYHSLRQPYA 659
QY 89 -P--D-----V--V-----Q--T--R-----Q-----Q--AY 99
Db 660 TNECKONNGCEQCVLSHRTDNDGLGRCKCTFGFOLDTDERHCIAVQNFIFSSQVAI 719
QY 100 R-----D-K-----L-AQ-----Q-----Q--A 108
Db 720 RGIPFTLSTQEDVMVPVSGNPSFFVGIDFDAQDSTIFFSDMSKMFQKIDGTGREILA 779
QY 109 A-----A-----A-----A-----A-----A-- 112
Db 780 ANRVENVESLAFDWISKNLWTDSHYKSI SVMRLADKTRTVVQYLLNPRSVVHPFAGY 839
QY 113 -----A-----A-----A-----A-----A-----S-- 119
Db 840 LFPTDWRPAKIMRAWSGDGSHLLPVINTILGWENGLAIDWAASRLYVWDAYDFKIEHSTF 899
QY 120 -----Q-----N-----GA--H-----T----- 120
Db 900 DGLDRRLGHIEQWTHPFGLAIFGEHLFFTDWRLGAIIRVRKADGEMTVIRSGIAYILH 959
QY 121 -----Q--GS--A-----KNG-----EN-----TA--N-----G-- 133
Db 960 LKSYDVNIQTGSNACNQPHTPNGDCSHFCPPVNFQVFCPCPYGMRLASNHLTCEGDPTN 1019
QY 134 E-----E-----N-----N-----GA--H-----T----- 140
Db 1020 EPTEQCGLFFPKCKGRCPVNYVLCGDVDDCHDNDSEQLCGTLNNTCSSAFTCGHGE 1079
QY 141 I-A-----N-----N-----H-----T--DME-----V-----DG-D 154
Db 1080 IPAHRCDKENDCVGSDSEHNCPTHAPASCLDTQYTCNHNQCISKWVCDTNDNDCGDS 1139
QY 155 -----VEI--P-----PN-----R-----K-A----- 162
Db 1140 EKNCTETCOPQFNCNHRICDLSFVGCDKDCVDGSDVGCVLNCTASQFKCASGDK 1199
QY 163 -----V-----VL-----R--G--H-----E----- 169
Db 1200 CIGVTRKCDGVFCDSDNSDEAGCTPRPGMCHSDEFOQSDGICIFNFEWCDGHPDCLYG 1259
QY 170 S-E-----V-----F-----IC--AW-----N-----P----- 179
Db 1260 SDEHNACVPKTCPSYFHCNDGNCIHRAWLCDRDNDGDMSDKDCPTQPPRCPFSQWQC 1319
QY 180 -----V-----V-----SD----- 182
Db 1320 LGHNICVNLVUCDGFDCPNGTDESPLNGNSCSDPFGGCTHECVQEPFGAKCLPLGP 1379
QY 183 LLA-----S-----S-----G----- 189
Db 1380 LLANDSKTCEDIDECIDILGSCSHQCYNMGRSPRSCDGTGYMLESDRGTCKVTASESILL 1439
QY 190 -----DS--T-----A-----RI-----W-----N----- 197
Db 1440 VASQNKIIADSVTSQVHNISYSLVENGSIYVAVDFDSISGRIFWSDATQGTWSAFQNGTD 1499
QY 198 -----LSB-----N--S--T-----S--GS-----T--Q--LVL 211
Db 1500 RRVFDSIILTBETIAIDWGRNLYWTDYALETIEVSKIDGSHRTVLSKNLTNPRGLAL 1559
QY 212 --R-----H----- 213
Db 1560 DPRMNEHLLFWSDGHHPRIERASMDGSMRTVIVQDKIFWPCGLTIDYPNRLLYFWD Syl 1619
QY 214 -----C-----IR-----E-----GG----- 219
Db 1620 DYMDFCDYNGHRRQVIASDLIIRHPYALTFLFEDSVYVTDTRATRRVRANKWHGQSVV 1679

QY 220 -----Q----- 220
Db 1680 MYNIQWPLGIVAVHPKQPNVNPCAFSRCSHCLLSSQGPHPFYSCVPSGWSLSPDLLN 1739
QY 221 -----D-----V-----P--SN--K-----DV----- 228
Db 1740 CLRDDQPFLLITVRAHQIIFGISLNPEVKSNDAWPIAGIQNGLDVEFPDABQYIYWVENPG 1799
QY 229 -----T-----S-----LDW-----N--S--E--G-----T 238
Db 1800 EHRVKTGDNRTVFAISIMVGPMSMLALDMSIRNLYSTNPTQSTIEVTLHGDIRYRKT 1859
QY 239 LLA-----T-----G--S-----Y-----DG-----F----- 248
Db 1860 LIANDGTALGVGPIGITVDPARGKLYWSDQGTDSGVPAKIASANMDGTSVKTLFTGNLE 1919
QY 249 -----A-----R--I-----W-----T----- 253
Db 1920 HLECVTLDIEEQKLYWAVTGRGVIERNVDGTDRMILVHOLSHPWGIAVHDSFLYTTDEQ 1979
QY 254 -----K--D--G--N--L-----ASTL--G--O--H-----K----- 266
Db 1980 YEVIERVKATGANKIVLRDNPVRLGLQVYHRRNAEBSNGCSNNMACQICLPVPGG 2039
QY 267 -----G-----P-----I--FALK----- 273
Db 2040 LFSCACATGFKLPNDRSCSPYNSFIVVSMLSAIRGFSLESDHSETWVPVAGQGRNALH 2099
QY 274 -----W-----N-----K-----K-----GN-- 279
Db 2100 VDVSSGFIYWCDFSSVASDRAIRRIKPDGSSLMNIVTHGIGENGVRGIAVDWVAGNL 2159
QY 280 ----- 280
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTDMPRHIVVDPKNRYLFWADYQORPKIERSF 2219
QY 281 -----I-----L-----S--A-----G----- 285
Db 2220 LDCNRTVLVSEGIPTVPRGLAVDRSDGYVYVWDDSLDIARIRINGENSEVIRYGSRYPT 2279
QY 286 -----VD-----K--T--TII-----W-----D----- 294
Db 2280 PYGITVFENSIWVDRNLKKIFQASKEPENTEPPTVIRDNNINWLRDVTIFDKQVQPRSPA 2339
QY 295 -----A-----HT-----G-----E-----A----- 300
Db 2340 EVNNPCLNNGCCHLCPALPGLHTPKCDCAFGLTQSDGKNCAISTENFLIFALNSLR 2399
QY 301 -----K-----Q-----Q--F-----P-- 305
Db 2400 SLHLDPENHSPPQTINVERTVMSLDYDSVSDRIYFTQNLASGVQISYATLSSGIHTPT 2459
QY 306 -----F----- 306
Db 2460 VIASGIGTAGIAFDWITRRIYVSYDYNQMINSMABDGSNRTVIARVPKPRAIVLDPQCG 2519
QY 307 -----H-----S--AP--ALD----- 313
Db 2520 YLYWADWDTHAKIERATLGNFRPVIWSSLVMPSGLTLDYEEDLLYWVDASLQRIERST 2579
QY 314 -----VD-----W-----Q-----S--N----- 319
Db 2580 LTGVDEVIWNAVHAFGLTYQYIYWTDLTYQRIYRANKYDGSQIANTNLLSQPRG 2639
QY 320 -NT-----F----- 322
Db 2640 INTVWKNQKQCNNPCEQFNGGCHICAPGPNGAECQCPHEGWNWYLANNRKHCIVDNGER 2699
QY 323 --AS--CS-----TD--M--C--IH-----V----- 333
Db 2700 CGASSFTCSNGRCISBEWKCDNDNDGSDGDEMESVCALHTCSPTAFTCANGRCVQYSYR 2759

334 CKL-----G--QD-----R-P--I-----342
2760 CDYNDGSGDEAGCLFRDCNATTFMCMNRRCIPREFICNGVCHDNNTSDEKCPD 2819
343 -----K-----T-----FQ-----346
2820 RTCOSYTKCHNSNICIPRYLVCDGNDGSDENPTYCTTHTCSSSEFQASGRICIPQ 2879
347 -----GH--T-----349
2880 HWYCDQBTDCFDASDEPASCCHERTCLADEFKDGGRCIPSEWICDGDNDGDMSEDK 2939
350 -----N--E--VNA-----IK--W-----357
2940 RHOCQNCQSDSFLCVNDRPRRCIPQSWVCDGVDCTDGYDENQNCRTTRTCSENEFT 2999
358 -----357
3000 CGYGLCIPKIFCRDRHNDGDSYDERGCLYQTCQOQFTCQNGRCISKTFVCDENDCGD 3059
358 --D-----PT-----G-----N-----L-----363
3060 GSDLMHLCHTPEPTCPPEBFKCDNGRCIEEMKLCNHLDDCLDSDSKGCGINECHDPSI 3119
364 -----LAS--CS-----D-D--M-----371
3120 SGCHNCTDLTSPYCSRGYKLMKRYCVDIDECTEMPVCSQKCNVIGSYICKCA 3179
372 -----371
3180 PGLYREPDGKTRQNSNIEPLIFSNRYLRLNTIDGYFYSLILEGLDNVVALDFRVEK 3239
372 -----T-----L--K--I-----W-----S--377
3240 RLYWIDTQROVIERMFLNKTNETIINHRLPAESLAVDWVRKLYLWDLARLDGLFVSD 3299
378 -----M-KO--D--N--C-----VHDLOQ-----H-----NK--392
3300 NGHRRMAQCHVDANNTFCFDPNPRGLAH--POGYLYWADWGHRAVIGRVGMDGTNKS 3357
393 -----E-----I-YT-----IK-----398
3358 VIISTKLEWNGITIDYNDLLYADAHGLEYEVSDELEGHHRHTVYDGPALPHPAITIFE 3417
399 -----W-----S-P--T--G--403
3418 DTIYTDWNTREVBKNGKIDGNSNRQTLVNTTHRPFDIHVYHPYRQPIVSNPCGTNGGCS 3477
404 -----PG-----T-----NN-----408
3478 HLCLIKPGGKFTCECPDDFTLQISGTYCMPMCSSTQFLCANNEKCIPIWKKCDGQKD 3537
409 -----P-----NA--N-----LML-----AS 417
3538 CSDGSELALCPORFRLGQFCSDGNGTSPQTLCAHQNGCDGSDERL--LCENHHCD 3596
418 -----A-----S-----F-----DS-----T-----VR-----L-----W--DV 429
3597 NEWQCANRKCIPESWQCTFDNDCEDNDESDSHCASRTCRPGQFRANGRCIPOAWKCDV 3656
430 --DRG-----I--CI--H-----T-L-----438
3657 DNDGDHSDPEIEBCMSAHLCDNFTEPSCNTYRCIPKVAWCVNGVDDCDRNSDEQCEE 3716
439 -T-----K--H-----Q-----E--P-----V-----445
3717 RTCHVPGDFRCKNHHCIPLRWQCDQNDGSDNSDEENCAPRECTESEFRVCNOQCIPSRW 3776
446 -----Y-----S--V-----A-----F 450
3777 ICDHYNDGDNDSRDERCMRTCHPBYFOCTSGHCVHSELKCDGSDADCLDASDEADCTRF 3836
451 SPDGRYL--AS-----GS-----F-----462

3837 -PDGAYCATMFECKNKHVCIPPYWKCDGDDDCGDSDEELHCLDLPVCPNSPFRCDNNR 3895
463 -----D-----K-----CV-H--I-----468
3896 CIYSHEVCNGVDDCGDGTDETEHCRKPTPKCTEYKCGNGHCIPHDNVCDADDCGD 3955
469 W-----N-TQ-----V-----CL-----475
3956 WSDLELCNKGKERTCAENICEQCTQLEGGFICSCTAGFETNVFDRISCLDINECEQFG 4015
476 -----H-----476
4016 TCPQCRNTKSGYECVADGFTMSDRPKRCAEGSPLLLLPDNRIRKYNLSERFS 4075
477 -YL-----480
4076 EYLODEEYIQAVDYDWDPKDIGLVVYTVRGEGRFGAIKRAYIPNFGSRNLLVQEV 4135
481 -----Q-----V-----L-----L-----N--485
4136 LKLKYMQPDGIADVWGRHIYSDVKNKRIEVAKLDRYKWLITDLDQPAIAVNP 4195
486 LG-----R-----SI-----490
4196 LGLMFWTDWKEPKIESAMWNGEDRNILVFDLGMPTGLSIDYLNNDRIYWSDFKEDVIE 4255
491 -----490
4256 TIKYDGTDRRTAKEMNPYSIDIFEDQLYWISKEGEVWKQNGKFGQKKEKTLVNPWL 4315
491 -----CL-----Y-----T-----494
4316 TVRIFPHQLRYNKSVPNLCKQICSHLCLLRPGYSACPGSSFTIEGSTTECDAAIELP 4375
495 -LP-----H-----497
4376 NLPPCRCHMGNCYFDTDLPRCKPGSGYTGKCEMAFSKGISPGTTAVALLTILLIV 4435
498 -----H-----LV-----V-----I-----502
4436 VIGALAIAGFFHVRRTGSLLPALPKPLSLSSLVKPSSENGVTFRSGADLNMDIGVSGFG 4495
503 P-----LVA-----LI-E--LL-----VLK 514
4496 PETADRSAMSDSDFVWEMGKQPIIFENPMYARDSAV-K 4534

RESULT 6

US-08-652-877-88

: Sequence 88, Application US/08652877

: Patent No. 6187548

: GENERAL INFORMATION:

: APPLICANT: Akerstrom, Goran

: APPLICANT: Juhlin, Claes

: APPLICANT: Raak, Lars

: APPLICANT: Crumley, Gregg R.

: APPLICANT: Morse, Clarence C.

: APPLICANT: Murray, Edward M.

: APPLICANT: Hjalms, Goran

: TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments

: INVENTION OF: Thereof and DNA Encoding Same

: NUMBER OF SEQUENCES: 106

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Rhone-Poulenc Rorer Inc.

: STREET: 500 Arcola Rd., 3C43

: CITY: Collegeville

: STATE: PA

: COUNTRY: USA

: ZIP: 19426-0107

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: Macintosh

```

Query Match          70.1%; Score 2558.8; DB 3; Length 4655;
Best Local Similarity 9.5%; Pred. No. 5.3e-32;
Matches 430; Conservative 60; Mismatches 18; Indels 4032; Gaps 371;

QY      1 M-----S-----I-----S--SDE- 7
      1 |DRGPAAVACTLLALLVACLAPASQBCSDSAHFCGSGHCIPADWRCDGTWDCSDDADEI 60
Db      1 ----VN-----F-----L-----V-----
QY      8 ----VN-----F-----L-----V-----
      61 |CAVVTCCQGYFKCQSBGQCIPISSWVCDQDQDCDSDERQDCSQSTCSSHQITCSNGQC 120
Db      61 |CAVVTCCQGYFKCQSBGQCIPISSWVCDQDQDCDSDERQDCSQSTCSSHQITCSNGQC 120
QY      13 ---YR-----Y-----L-----Q-----E--S 19
      121 |PSEYRCDHVRDPCPDGADENDCQPTCEQLTCDNGACVNTSQKCDWKVCDKRDSSDEINCT 180
Db      121 |PSEYRCDHVRDPCPDGADENDCQPTCEQLTCDNGACVNTSQKCDWKVCDKRDSSDEINCT 180
QY      20 ---G-FS-----H-----S--A--F--T---F--G--I--- 30
      181 |EILHNEFSCNGECIPRAYVCDHNDQDGSDEHACNYPTCGGYQTCPSGRCIYQNWV 240
Db      181 |EILHNEFSCNGECIPRAYVCDHNDQDGSDEHACNYPTCGGYQTCPSGRCIYQNWV 240
QY      31 ---K-----S-----H-----IS-----Q--SNI 39
      241 |CDGEDCKNGDEDCGSESHVDHVKSCPREWSCSPESGRCISIKVCDGILDCPGREDEN- 299
Db      241 |CDGEDCKNGDEDCGSESHVDHVKSCPREWSCSPESGRCISIKVCDGILDCPGREDEN- 299
QY      40 N--G-----AL-----VP-----PAALI--S-----I--I-- 53
      300 |NTSGKYCSMTLCSALNCQYQCHETPYGGACFCPPGYIINHDSRTCVFEPDQCIWGLCD 359
Db      300 |NTSGKYCSMTLCSALNCQYQCHETPYGGACFCPPGYIINHDSRTCVFEPDQCIWGLCD 359
QY      54 QK-----G--L--QY-----VE 61
      360 |QKESRPRGRHLCHCEGYILLERGQYCKANDSFGEASIFISNGRDLICDIHGRSFRILVE 419
Db      360 |QKESRPRGRHLCHCEGYILLERGQYCKANDSFGEASIFISNGRDLICDIHGRSFRILVE 419
QY      62 ---A-----E--VSI-----N 67
      420 |SONRGVAVGAFHYHLQRFVFTWDTVQNKFVSVDINGLNIQVLNVSVETPENLAVDWVNN 479
Db      420 |SONRGVAVGAFHYHLQRFVFTWDTVQNKFVSVDINGLNIQVLNVSVETPENLAVDWVNN 479
QY      68 ---E-----DG--TL----- 72
      480 |KIYIVETKVNIDMWNLDDGSVRVTITENLGHPRGIAVDPVGYLFFSDWESLSGEKLE 539
Db      480 |KIYIVETKVNIDMWNLDDGSVRVTITENLGHPRGIAVDPVGYLFFSDWESLSGEKLE 539
QY      73 --F-DG--R-----P-----IE-----S 80
      540 |RAFMDGNSRKDLVKTKLGWPAGVTLDMISKRVYVWDSRFDYIETVTVDGIORKTWVHGS 599
Db      540 |RAFMDGNSRKDLVKTKLGWPAGVTLDMISKRVYVWDSRFDYIETVTVDGIORKTWVHGS 599

```

QY 220 -----Q----- 220
Db 1680 MYNIQWPLGIVAVHPKQPNVNPFCRSHCLISSQGPFPYSCVCPGSGWSLSPDLLN 1739
QY 221 -----D-----V-----P-----SN-----K-----DV----- 228
Db 1740 CLRDDQPLITVQHOIIFIGISLNPVKNSNDAMVPIAGIQNGLDVEFDDAEQYIYWENPG 1799
QY 229 -----T-----S-----LDW-----N-----S-----E-----G-----T 238
Db 1800 EIHVRKTDGTRTVPFASISVMGFSMMLALDWISRLYSTNPTQSTIEVLTLHGDIRYKT 1859
QY 239 LLA-----T-----G-----S-----Y-----DG-----F----- 248
Db 1860 LIANDGTALGVGPPIGITVDPARGKLYWSDQGTSDGVPAPAKIASANWDGTSVKTLFTGNLE 1919
QY 249 -----A-----R-----I-----W-----T----- 253
Db 1920 HLECVTLIDIEEQKLYWAVTGRGVIERNVDGTRDMILVHQLSHPWGLIAVHDSFLYYTDEQ 1979
QY 254 -----K-----D-----G-----N-----L-----ASTL-----G-----Q-----H-----K----- 266
Db 1980 YEVIERVDKATGANKIWLDRNVENLRLGLQVYHRRNAAESSNGCSNNMACQQICLPVPG 2039
QY 267 -----G-----P-----I-----FALK----- 273
Db 2040 LFSCACATGPKLPDNRSCSPYNSFIVVMSLAIRGFSLESDHSETWVPVAGQGRNALH 2099
QY 274 -----W-----N-----K-----GN----- 279
Db 2100 VDVSVSGFIYWCDFSSVASDANAIIRIKPDGSSLMNIVTHGIGENGVRGIAVDWVAGNL 2159
QY 280 -----I-----L-----S-----A-----G----- 285
Db 2160 YFTNAFVSETLIEVIRINTTYYRVLKVTVDMPRHIVVDPKNYLFWADYQRPKIERSF 2219
QY 281 -----I-----L-----S-----A-----G----- 285
Db 2220 LDCNRTVLSGIVTPRGIAVDRSDGYVWVDDSLDIIRIRINGENSEVIRYGSRYPT 2279
QY 286 -----VD-----K-----T-----TII-----W-----D----- 294
Db 2280 PYGITVFENSIIWDRNLKXIFQASKEPENTPTVIRDNINMLRDVTIFDKQVQPRSPA 2339
QY 295 -----A-----HT-----G-----E-----A----- 300
Db 2340 EVNNPCLNNGGCSHLCPALPLHTPKDCAFQTLQSGDKGNCAISTENFLIFALSNSLR 2399
QY 301 -----K-----Q-----F-----P----- 305
Db 2400 SLHLDPENHSPFQTIINVERTVMSLDYSDSDRIYFTQNLASGVQISYATLSSGIHTPT 2459
QY 306 -----F----- 306
Db 2460 VIASGIGTADGIAFDWITRIRIYSDYLNQMINMAEDGNSRNVIRIARVPRVPRVLDPCQ 2519
QY 307 -----H-----S-----AP-----ALD----- 313
Db 2520 YLYWADWDTHAKIERATLGGNFRVPIVNSLSVMSGLTLDYEEDLLYWDASLQRIERST 2579
QY 314 -----VD-----W-----Q-----S-----N----- 319
Db 2580 LTGVDRVINVAHAFGLTLYGYIYWDLTQYRIYRANKYDGGQIAMTTLNLSQPRG 2639
QY 320 -----NT-----F----- 322
Db 2640 INTVVKNQKQCNPNFCEQFNGGCSHICAPGNGAECQCPHEGNYLANNRKHCIVDNGER 2699
QY 323 -----AS-----CS-----TD-----M-----C-----IH-----V----- 333
Db 2700 CGASSFTCSNGRCIISEWKDNDNDGSGDEMESVCAIHTCSPTATCANGRCVQYSYR 2759
QY 334 CKL-----G-----QD-----R-----P-----I----- 342

Db 2760 CDYYNDGSGDEAGCLFRDCNATTEFCMNNRRCIPREFICNGVDNCHDNNNTSDEKNCPD 2819
QY 343 -----K-----T-----FQ----- 346
Db 2820 RTCQSGYTKCHNSNICIPRVYLCGNDGNDSDENPTYCTHTTCSSEFQOCASRCIPQ 2879
QY 347 -----GH-----T----- 349
Db 2880 HWYQOETDCFPASDEPASCSEHSEPTCLADEPKDGGRCIPSEWICDGDNDGDMSEDK 2939
QY 350 -----N-----E-----VNA-----IK-----W----- 357
Db 2940 RHQCQNQNSDSEFLCVNDRPPDRRCIPQSMVWCDGVDCTDGYDENQNCNTRTCSNEFT 2999
QY 358 ----- 357
Db 3000 CGYGLCIPKIPRCDRHNDGYSYDGRGCLYQTCQOQNTCQNGRCISKTFVCDENDCGD 3059
QY 358 -----D-----PT-----G-----N-----L----- 363
Db 3060 GSDELMHLCHTPEPTCPPEFKCDNGRCIEMMKLCNHLDDCLDSDSEKGCINECHDPSI 3119
QY 364 -----LAS-----CS-----D-----D-----M----- 371
Db 3120 SGCDHNCTDLTATFYCSCRPGYKLSMDKRTCDVIDECTEMPEFVCSQKCNVIGSYICKA 3179
QY 372 ----- 371
Db 3180 PGLYREPDKTCRONSNEPYLIFSNRYLRNLITIDGYFYSILLEGLDNVVALDFRVEK 3239
QY 372 -----T-----L-----K-----I-----W-----S----- 377
Db 3240 RLYWIDTQOVIERNFLNKTETIINHRLPAESLAVDWVSRKLYWLDARLDGLFVSDL 3299
QY 378 -----M-----KQ-----D-----N-----C-----VHDLQO-----H-----NK----- 392
Db 3300 NGGHRMLAQHCVDANNTECFDNPRGLAH--POYGYLYWADWGHRAIYGRVGMGTNKS 3357
QY 393 -----E-----I-----YT-----IK----- 398
Db 3358 VIISTKLEPNGITIDYNDLLYWADAHLYEYSDLEGHHRHTVYDGLPHFPFAITIE 3417
QY 399 -----W-----S-----P-----T-----G----- 403
Db 3418 DIIYWDNTRIVKEKNGYDGSNRQTLVNTHTRPDIHVHPYRQPIVSNPCGTNNGGCS 3477
QY 404 -----PG-----T-----NN----- 408
Db 3478 HLCLIXPGKGFTCECPDPTLQLSGSTYCMPCSSOTFLCANNKECIPIWKKDQKD 3537
QY 409 -----P-----NA-----N-----LML-----AS 417
Db 3538 CSDGSELALCPORFCRLGQFCQSGNCTSPQTLNQHONCPDGSDEDL--LCENHHDS 3596
QY 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-----DV 429
Db 3597 NEWQANKRCIPESWQCDTFNDCEDNSEDSDSHCASRTCRPGQFRCANGRCIPQAWKCDV 3656
QY 430 -----DRG-----I-----CI-----H-----T-----L----- 438
Db 3657 DNDGSHSDEPIBECMSSAHLCDNFTEFSCKTNYRCIPKAVCNVDDCDRDSDEQCEE 3716
QY 439 -----T-----K-----H-----Q-----E-----P-----V----- 445
Db 3717 RTCHPVGDPECKNHHCIPLRWQCDGNDGNDSENCAPRECTESEFRVCVQCIPSRW 3776
QY 446 -----Y-----S-----V-----A-----F 450
Db 3777 ICDHYNDGNSDERDCMRTCHPEYFQCTSGHGVHSELKCDGSADCLDASDEADCFTRF 3836
QY 451 SPGRYL--AS-----GS-----F----- 462

Db 3837 PDGAYQATMFECKNHVCIPYWKCDGDDCGSGDEELHCLDVPCNSPNRFRCDNNR 3895
Qy 463 -----D-----K-----CV-H-I----- 468
Db 3896 CIYSHEVCNGVDCGDTDEBEHCRKPTPKPCTEYKCGNGHCIPHNDVNCDDADDGCD 3955
Qy 469 W-----N-TQ-----V-----CL----- 475
Db 3956 WSDLGCKGKERTCAENICQNTQNEGGFICSCCTAGFTNVFRTSCLDINECEQFG 4015
Qy 476 -----H----- 476
Db 4016 TCQHCNTKGSYECVADGFTSMRDPKRCACAGSSPLLLPDNVRIRKYNLSRERS 4075
Qy 477 YL-----Q-----N-G----- 480
Db 4076 EYLQDEYIQAVDYDMDPEDIGLSVVYTVVRGSRFGAIKRAYINFESGRNLLQVEVD 4135
Qy 481 -----Q-----V-----L-----N----- 485
Db 4136 LKLKYYMQDGIADVWVGRHRYWSDVKNKRIEVAKLDGRYRWLSTDLDPQAAIAVNP 4195
Qy 486 LG-----R-----SI----- 490
Db 4196 LGLMFTWDMCKEPKESAMNGEDRNILVFEDLGWPTGLSIDYLNNDRIYWSDFKEDVIE 4255
Qy 491 ----- 490
Db 4256 TIKYDGTDRRIAKEAWNPYSLDIPEDOLYISKEGEVWKQKFGGKKEKTLVNPWL 4315
Qy 491 -----CL-----Y-----T----- 494
Db 4316 TVRIFHQLRYNKSVPNLCKQICHLRLPGGYSACPOGSGFIEGSTECDAAIPLI 4375
Qy 495 -LP-----H----- 497
Db 4376 NLPFPCRMHGNCPYFDTDLPKCKPFGYTGKCEMAFSKGISPGTTAVALLTILLIV 4435
Qy 498 -----H-----LV-----V-----I----- 502
Db 4436 VICALAIGFFHYRRTGSLLPALPKLPSSLVKPSENGVTFPSGADLNMDIGVSGFG 4495
Qy 503 P-----LVA-----LI-E--LL-----VLK 514
Db 4496 PETADRSMAMSEDFVWEMGKQPIIPENPMYSARSAV-K 4534

RESULT 7

US-08-652-877-90
; Sequence 90, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Rask, Lars
; APPLICANT: Crumley, Gregg R.
; APPLICANT: Morse, Clarence C.
; APPLICANT: Murray, Edward M.
; APPLICANT: Hjalms, Goran
; TITLE OF INVENTION: Human Calcium Sensor Protein; Fragments
; TITLE OF INVENTION: Thereof and DNA Encoding Same
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd., 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426-0107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.5.1

SOFTWARE: Word 6.0 (Patentin)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/652,877
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: PCT/US95/15203
APPLICATION NUMBER: PCT/US95/15203
FILING DATE: 22-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,836
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/487,314
FILING DATE: 07-JUNE-1995
ATTORNEY/AGENT INFORMATION:
NAME: Savitzky, Martin
REGISTRATION NUMBER: 29,699
REFERENCE/DOCKET NUMBER: A1355E-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-454-3816
TELEFAX: 610-454-3808
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 4655 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-652-877-90

Query Match 70.1%; Score 2558.8; DB 3; Length 4655;
Best Local Similarity 9.5%; Pred. No. 5.3e-32;
Matches 430; Conservative 60; Mismatches 18; Indels 4032; Gaps 371;

Qy 1 M-----S-----I-----S-----S-DE- 7
Db 1 MDGPAVACTLLALVACLAPASGQECDSAHFRCSGHCI PADWRCGTGDKCDDADEI 60
Qy 8 ---VN---F---L---V----- 12
Db 61 GCNAVTCQGYKQSEGGCIPSSWVCDQDQDCDDGSDERQDCSQSTCSHQITCSNGQC 120
Qy 13 ---YR-----Y-----L-----Q-----E--S 19
Db 121 IPSEYRCHDVRCPDGADENDCQYPTCEQLTCDNGACYNTSQKCDWKVDCRSDSEINCT 180
Qy 20 -----G-FS-----H-----S-A-F-T-----F---G-I----- 30
Db 181 EICLHNEFSCNGECIPRAYVCDHNDCCQDGSDEHACNYPTCGYQFTCPSGRCIYQNV 240
Qy 31 -----K-----S-----H-----IS-----Q--SNI 39
Db 241 CDGEDDCKNGDEDCESGPHDVHKCSPREWSCPESGRCISYKVCDCGILDCPREDEN- 299
Qy 40 N--G-----AL-----VP-----PAALI-----S-----I--I-- 53
Db 300 NTSTGKCSMTLCSALNCOYQCHETPYGACFCPPGYIINHDSRTCTCVFDDCQIWIICD 359
Qy 54 QK-----G-L---QY-----VE 61
Db 360 QKCESRPGRHLCHEEGYILERGQYKANDSPGEASIIFNSGRDLLIGDINGHSRFLVE 419
Qy 62 -----A-----E--VSI-----N 67
Db 420 SQNRGAVGVAFYHQLQRFVFTDVTQNVKVPFVSDINGLNIOEVLNVSVETPENLAVDWNV 479
Qy 68 -----E-----DG-----TL----- 72
Db 480 KIYLVETKVNRI DMVNLDSYRVTLITENLHGRGIAVDPTVGYLFFSDWESLSGPKLE 539
Qy 73 --P-DG--R-----P-----IE-----S 80
Db 540 RAFMDGSRNKKDLVKTGLGWPAGVTGLDMISKRVYVWDSRFDYIETVTYDGIQRTVTVHGS 599

Db 2760 CDYNDGCGSDEAGCLFRDCNATTFBMCNRRRCIPREFICNGVDNCHDNTSDEKXCPD 2819
QY 343 ---K---T---FQ---346
Db 2820 RTCQGYTKHNSNICIPRYVLCDGNDGCDNSDENPTYCTHTCSSEFQASGRICIPQ 2879
QY 347 ---GH---T---349
Db 2880 HWYCDQETCFDASDEPASCGHSERTCLADEFKDGGRCIPSEWICDGDNDGMSDEDK 2939
QY 350 ---N---E---VNA---IK---W---357
Db 2940 RHOCQONQSDSEFLCVNDRPPRRRCIPQSDWVCDGVDCTGDYDENQNCNTRTCSNEFT 2999
QY 358 ---357
Db 3000 CGYGLCIPKIFPCDRHNDGDSYDERGCLYQTCQOQFTCQNGRCISKTFVCDENDDNGD 3059
QY 358 ---D---PT---G---N---L---363
Db 3060 GSDELMHLCHTPEPTCPPEHFKDNGRCIEMMKLCNHLDDCLDSDSKGCGINECHDPSI 3119
QY 364 ---LAS---CS---D---D---M---371
Db 3120 SGCDHNCITDLTSFYCSRCRPGYKMSDKRKTCDVIDECTEMPFVCSQKCNVIGSYICKCA 3179
QY 372 ---371
Db 3180 PGYLRPDGKTCQNSNIEPYLIFSNRYLRNUTIDGYFVSLILEGLDNNVALDFRVEK 3239
QY 372 ---T---L---K---I---W---S---377
Db 3240 RLYWIDTQROVIERMFLNKTNETIINHLPAEASLAVDWVSRKLYWLDARLDGLFVSDL 3299
QY 378 ---M---KQ---D---N---C---VHDLOQ---H---NK---392
Db 3300 NGHREMLAQHCVDANNTCFDNPRLALH--PQYGYLWADWGHRAIYGRVGMGTNKS 3357
QY 393 ---E---I---YT---IK---398
Db 3358 VIISTKLEWNGITIDYNDLLYWAHAHLGYIEYSDLGHRHTVVDGALPHFPALTIFE 3417
QY 399 ---W---S---P---T---G---403
Db 3418 DTIYTDNTRTVEKNGYDGSNRQTLVNTHTRPDIHVHPYRQPIVSNPCGTNNGCS 3477
QY 404 ---PG---T---NN---408
Db 3478 HLCILKPGKGFTCECPDDPRTLQLSGTYCMPMCSSTOPLCANNEKICPIWVKCDGQXD 3537
QY 409 ---P---NA---N---LML---AS 417
Db 3538 CSDGSDELALCPORFRLGQFCQSGNGCTSPTQLCNAHNCNCPDGSDEDL--LCENHHCD 3596
QY 418 ---A---S---F---DS---T---VR---L---W---DV 429
Db 3597 NEWQCANRCIPESWQCDTFNDCEDESDSSHCASRTCPGQFRANGRCIPQAKCDV 3656
QY 430 ---DRG---I---CI---H---T---L---438
Db 3657 DNDGDSDEPIBECMSAHLCDNFTFESCKTYRCIPKWAQVGVDDCDRDSDEQGBE 3716
QY 439 ---T---K---H---Q---E---P---V---445
Db 3717 RTCHPVDFRCNHHCIPLRWQCDGNDGCDNSDEENCAPRECTESEFCVNOQCIPSRW 3776
QY 446 ---Y---S---V---A---F 450
Db 3777 ICDHYNDGCDNSDERCEMRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCTPF 3836
QY 451 SPDGRYL-AS-----GS-----F-----462
Db 3837 -PDGAYCOATMFBCKNHVCIPPYWKCDGDDDCDGDGDEELHCLDVPNCSPNFRFRCDNNR 3895

QY 463 ---D---K---CV-H-I---468
Db 3896 CIYSHEVCNVDGCGDTDETEHCRKPTPKPCTEYKCGNGHCIPHDNVCDADDDCGD 3955
QY 469 W---N---TO---V---CL---475
Db 3956 WSDLGCKNGKERTCAENICEONCTQNBGGFTCSCTAGFETNVFDRTSCLDINECEQFG 4015
QY 476 ---H---476
Db 4016 TPOHCRNTKGYECVACDGTSMDSRPGKRCAAEGSSPLLLPDNVIRKYNLSRFRS 4075
QY 477 -YL---N---G---480
Db 4076 EYLQDEEYQAVDYDWDPEDIGLSVVVYTVRGESRFGAKRAYIFNFESGRNLLVQEV 4135
QY 481 ---Q---V---L---L---N---485
Db 4136 LKLKYMVQPDGIAVDWVGRHIYMSDVKNKRIEVAKLDGRYKWLISTDLDPAAIAVNP 4195
QY 486 LG---R---SI---490
Db 4196 LGLMFTWDGKBPKELESAMNGEDRNILVPDLGWPTGLSIDYLNNDRIYWSDFKEDVIE 4255
QY 491 ---490
Db 4256 TIKYDGTDRRVIKAEANPYSLDIPEDQLYISKEKEVWKQKFGQKKEKTLVNPWL 4315
QY 491 ---CL---Y---T---494
Db 4316 TQVRIFHQLRYNKSVPNLCKQICSHLCRLPGGYSCAPQGSFIEGSTTECDAIELPI 4375
QY 495 -LP---H---497
Db 4376 NLPPRCRWHGNCYFDETDLPKCKPCPSGYTKYCEMAFSKGISPGTTAVALLTILIV 4435
QY 498 ---H---LV---V---I---502
Db 4436 VIGALAIAGFFHYRTGSLPALPKPLPSLSLVKPSSENGVTFRSGADLNMDIGVSGFG 4495
QY 503 P---LVA---LI-E--LL---VLK 514
Db 4496 PETADRSWAMGSEDFVWEMGKQPIIFENPMYSARDSAV-K 4534

RESULT 8

US-08-652-877-86

; Sequence 86, Application US/08652877

; Patent No. 6187548

; GENERAL INFORMATION:

; APPLICANT: Akerstrom, Goran

; APPLICANT: Juhlin, Claes

; APPLICANT: Raak, Lars

; APPLICANT: Crumley, Gregg R.

; APPLICANT: Morse, Clarence C.

; APPLICANT: Murray, Edward M.

; APPLICANT: Hjalms, Goran

; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments

; TITLE OF INVENTION: Thereof and DNA Encoding Same

; NUMBER OF SEQUENCES: 106

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Rhone-Poulenc Rorer Inc.

; STREET: 500 Arcola Rd., 3C43

; CITY: Collegeville

; STATE: PA

; COUNTRY: USA

; ZIP: 19426-0107

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Macintosh

; OPERATING SYSTEM: System 7.5.1

; SOFTWARE: Word 6.0 (patentin)

Db 1680 MYNIQWLGIVAVHPKQPNVNCAPFRCRSHCLJSSQGHFYSCVPCSGWSLSPDLLN 1739
Qy 221 ---D---V---P---SN---K---DV--- 228
Db 1740 CLRDDQPLITVROHIIIFGISLNPVKSNDAMVPIAGIQNGLDVEFDADQYIYWENPG 1799
Qy 229 ---T---S---LDW---N---S---B---G---T 238
Db 1800 EIHVRKTDGNTVFASISMVGPSSMLALDWISRNLYSTNPTQSIETVLTHGDIRYKRT 1859
Qy 239 LLA---T---G---S---Y---DG---F--- 248
Db 1860 LIANDGTALGVGPFPIGITVDPARGKLYWSDQSDGVPKAKIASANMDGTSVKTLFTGNLE 1919
Qy 249 ---A---R---I---W---T--- 253
Db 1920 HLECVTLIDEEQKLYWAVTCRGVIERGNVDGTRMILVHQLSHPWGIAVHDSFLYTDEQ 1979
Qy 254 ---K---D---G---N---L---ASTL---G---Q---H--- 266
Db 1980 YEVIERVDRATGANKIVLRDNVPLRGLQVYHRRNAAESSNGCSNNMNACQOICLPVPGG 2039
Qy 267 ---G---P---I---FALK--- 273
Db 2040 LFSCACATGKLPNDRNSCPYNSFIWVMSLSAIRGFSLESDHSETMVPVAGQGRNALH 2099
Qy 274 ---W---N---K---GN--- 279
Db 2100 VDVVSSGFIWYCDSSVASDRAIRIKPDGSSLMNIVTHGIGENGVRGIAVDWVAGNL 2159
Qy 280 ---I---L---S---A---GN---F 280
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVDMPRHIVDPKRYLFWADYQRPKIERSF 2219
Qy 281 ---VD---K---T---TII---W---D--- 294
Db 2280 PYGITVFENSIWDRNLKFIPOASKEPENTPEPTVIRDNINMLRDTVFDKQVQRPSPA 2339
Qy 295 ---A---HT---G---E---A--- 300
Db 2340 EVNNPCLENNGCSHLCPALPGLHPTKCDKAFGLTQSDGKNCAISTENFLIFALSNSLR 2399
Qy 301 ---K---Q---Q---F---P--- 305
Db 2400 SLHLDPENHSPPTQINVERTVMSLDYSDRIYFTQNLASGVQISYATLSSGIHTPT 2459
Qy 306 ---F--- 306
Db 2460 VIASGIGTAGIADWITRIIYSDYLNQMINSMASDGNRTVIARVPKRAIIVLDPQCG 2519
Qy 307 ---H---S---AP---ALD--- 313
Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSLVMPSGTLIDYEDLLYWDASLQRIERT 2579
Qy 314 ---VD---W---Q---S---N--- 319
Db 2580 LTGVDRREVINAHAVHAFGLTLYQYIYWTDLTYQRIYRANKYDGSQIAMTTLNLSQPRG 2639
Qy 320 ---NT---F--- 322
Db 2640 INTVKNQKQCNPCQENGGCSHTCAPGNGAECQCQPHGNYLANNRKHCIVDNGER 2699
Qy 323 ---AS---CS---TD---M---C---IH---V--- 333
Db 2700 CGASSFTCSNGRCISBEWKDNDNDGDSDEMESVCALHTCSPTAFTCANGRCVQYSYR 2759
Qy 334 CKL---G---OD---R---P---I--- 342
Db 2760 CDYINDGDSDEAGCLFRDCNATTEFCMNRRCIPREFICNGVDNDNDNTSDEKNCPD 2819

Qy 343 ---K---T---FQ--- 346
Db 2820 RTCQSGYTKCHNSNICIPRVYLCDGNDCCGNSDENPTYCTTHTCSSEFQCTSGRCIPIQ 2879
Qy 347 ---GH---T--- 349
Db 2880 HMWCDQETDCFADSPASCGHSERTCLADEFKCDGRCIPSEWICDGDNDGDMSDDEK 2939
Qy 350 ---N---E---VNA---IK---W--- 357
Db 2940 RHQONQNSDSEFLCNDVRPPDRRCIPQSWVWCGDVDCDGDYDENQNCNTRTTCSENEFT 2999
Qy 358 --- 357
Db 3000 CGYGLCIPKIFRCDRHNDGDSYSDERGCLYQTCQOQNFQCONGRCSKTFVCDENDDCGD 3059
Qy 358 ---D---PT---G---N---L--- 363
Db 3060 GSDELMHLCHTPEPTCPPEFKCDNGRCIEMMKLCNHLDDCLNDSDEKGGCINECHDPSI 3119
Qy 364 ---LAS---CS---D---D---M--- 371
Db 3120 SGCDHNCTDLTSFYCSRPGYKLMGDKRTCVDIDECTEMPFVCSQKCNVIGSYICKCA 3179
Qy 372 --- 371
Db 3180 PGYLRBPDGKTCRONSIEPYLIFSNRYLRLNLTIDGYFSLILEGLDNNVALDFORVEK 3239
Qy 372 ---T---L---K---I---W---S--- 377
Db 3240 RLYWIDTQOVIERMFLNKTNETIINHRLPAAESLAVDWVSRKLYWLADARLQGLFVSDL 3299
Qy 378 ---M---KQ---D---N---C---VHDLOO---H---NK--- 392
Db 3300 NGGHRMLAQHCVDANNTFCFDPNPRGLALH---PQGYLYWADMGHRAIYGRVMDGTNKS 3357
Qy 393 ---E---I---YT---IK--- 398
Db 3358 VIISTKLEWPGITIDYTDNLLYWADAHLYEYSLEGGHRRHTVYDGLPHEPAITIFE 3417
Qy 399 ---W---S---P---T---G--- 403
Db 3418 DTIYTDWNTRTVEKNGKYDGNRQTLVNTTHRPFDIHVYHPYRQPIVSNPCGTNNGCS 3477
Qy 404 ---PG---T---NN--- 408
Db 3478 HLCLIKPGKGFTCECPDDFTLQLSGSTYCMPCSSSTQFLCANNEKCIPIWVKCDGQKD 3537
Qy 409 ---P---NA---N---LML---AS 417
Db 3538 CSDGSELAALCPQRCRLQFQSGDGNCTSPOTLCNAHQNCPPDGSDEDLJ---LCENHHCS 3596
Qy 418 ---A---S---F---DS---T---VR---L---W---DV 429
Db 3597 NEWQCANKECIPESWOCDFNDCEDESDSHCSARTCRPGCFRANGRCIPQANKCDV 3656
Qy 430 ---DRG---I---CI---H---T---L--- 438
Db 3657 DNDGDSHDEPIEECMSSAHLCDNFTEFSCKTNYRCIPKWAVCNGVDDCDRNSDEQGBE 3716
Qy 439 ---T---K---H---Q---E---P---V--- 445
Db 3717 RTCHPVGDPRKNHHCIPURWOCQDQNDGDSDEENCAPRECTESEFCVNOQCIPSRW 3776
Qy 446 ---Y---S---V---A---F 450
Db 3777 ICDHYNDGDSNDRDCERMRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCTPRF 3836
Qy 451 SPDGRYL---AS---GS---F--- 462
Db 3837 -PDGAYCOATMECKNHVICIPPYWKDGDGDDCGDGSDEBELHCLDVPNCSPNFRCDNNR 3895

QY	463	-----D-----K-----CV-H-I-----	468	Db	1	MERSGARQTFLLALLGATLRARAAAGYPRFSPFFFLCTHHGLEGEGQGEVLISLH	60
Db	3896	CIYSHEVCNGVDCGDTDETEHCKRTPKPCTEYKCGNGHCIPHNVCDDADDGCD	3955	QY	9	---N---F---LV---VRY---LQ---ES---CF---S-H-23	23
QY	469	W-----N-TQ-----V-----CL-----	475	Db	61	IAGNPTYYVPGQYHVHTISTTFFDGLLVGL-YTSTVSQASOSIGSSAFGFGIMSDHQ	119
Db	3956	WSDGLGKNGKERTCAENICEQNTQLNEGFFICSTAGFETVDPRTCLDINECEQFG	4015	QY	24	---S-A---F---T---F---G-I-K---31	31
QY	476	-----H-----	476	Db	120	FGNQFMCVWASHVSHLPTTNLSFIWAPPAGTCGVNFMATATHRGQVIFKDALAAQOLCE	179
Db	4016	TCPOHCRNTKGYECVACDGTSMRPGKRCACAGSSPLLLPDNVRIRKYNLSRERS	4075	QY	32	---S---H---I---S---Q-SNI---N---GAL---43	43
QY	477	YL-----N-G-----	480	Db	180	QCAPTDTVTHPLABIHSDSIILRDDPDSYHQLQNPNIWECNCETGEOCGAIMHGNA	239
Db	4076	EYLODEYIOAVDYDMDPDXDIGLSVVYTVRSGSRPGAIRKRAYIENFESGRNVLQEVVD	4135	QY	44	V---P-P---A-AL---I---SI---IQ---54	54
QY	481	-----Q-----V-----L-----N-----	485	Db	240	VTFCEPYGRELITGLNTTASVLQFSGSCSRFSYSDPSIIIVLYAKNSADMIQLEK	299
Db	4136	LKLYWQPDGIAVDWGRHIYHSDVKNKRIEVAKLDGRYRKWLISTDLDPAAIAVNP	4195	QY	55	---KG---LQY---V-EA-E---VSIN---67	67
QY	486	LG-----R-----SI-----	490	Db	300	IRAPSNVSTIIHILYLPEDAKGENVQFQWQENLRNVEYVACWALDNILI-INS	358
Db	4196	LGLMFTDMGKPKXBSAMNGEDRNILVPEDLGPWTGLSIDYLNDRYIWSDFKEDVIE	4255	QY	68	---ED---GT-IF---DG---R---76	76
QY	491	-----	490	Db	359	VLEDSLDPVDTGNWLFPPGATVKHSCQSDGNSIYPHNEGSEFNFATTRDVLSTEDIOE	418
Db	4256	TIKYDGTDRRVIKAEAMNPYSLDIFEDQLTWISKEGEVWKQKFGGKKEKTLVNPWL	4315	QY	77	---P---IES-LSLI---DA---86	86
QY	491	-----CL-----Y-----T-----	494	Db	419	QWSEFPESOPTGWDVLGAVIGTECGTIESGLSMVFLKDGKRLCTPMSMTTGYGNLRFYF	478
Db	4316	TQVRIFHQLRYKSVNPLCKQICSHLCLLRPGGYSCACPOGSGFIEGTTCCDAAIELPI	4375	QY	87	VM---P---D---D---D---V---QT---94	94
QY	495	LP-----H-----	497	Db	479	VNGGICDPCNSHENDIILYAKIEGRKEHITLDTLSYSSYKVPSPVSVVINPELOTATPKF	538
Db	4376	NLPPPCRMHGNCYFDETDLPKCKPSPGYTKYCEMAFSGISPGTTAVAVLLTILLIV	4435	QY	95	---RQ---QAY-R---D---K-LAQ---Q---106	106
QY	498	-----H-----LV-----V-----I-----	502	Db	539	CLRKQKHOGHNRNVAVDFFHVLVPLPSTMHMIQPSINLGCQTHQPGMSVLEFSTNHG	598
Db	4436	VIGALATAGFHYRRTGSLPALPKPLSLSLVXPSENGVTFRSGADLNMDIGVSGFG	4495	QY	107	QA---A---AAA---AA---114	114
QY	503	P-----LVA-----LI-E-LL-----VLK 514		Db	599	RSWSLLHTECLPEICAGPLPHSTVSYSENYSYGNWNRITIPNAAALNTRINRWRGTPI	658
Db	4496	PETAIDRSMAMSEDFVMEMGQPIIFENPMYSARDSAV-K 4534		QY	115	---A---A---AASQ---Q-GSA 124	124
RESULT 9							
US-09-334-220-1							
; Sequence 1, Application US/09334220							
; Patent No. 6323177							
GENERAL INFORMATION:							
; APPLICANT: St. Jude's Children's Research Hospital							
; APPLICANT: Curran, Thomas							
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW							
; TITLE OF INVENTION: D'Arcangelo, Gabriella							
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND							
; TITLE OF INVENTION: THERAPIES							
; FILE REFERENCE: 2427/0F704							
; CURRENT APPLICATION NUMBER: US/09/334,220							
; CURRENT FILING DATE: 1999-06-16							
; NUMBER OF SEQ ID NOS: 5							
; SOFTWARE: FastSeq for Windows Version 3.0							
; SEQ ID NO 1							
; LENGTH: 3460							
; TYPE: PRT							
; ORGANISM: Homo sapien							
US-09-334-220-1							
Query Match 70.0%; Score 2554; DB 3; Length 3460;							
Best Local Similarity 11.3%; Pred. No. 1.3e-32;							
Matches 391; Conservative 95; Mismatches 14; Indels 2950; Gaps 357;							
QY	1	M-S-----I-----S-----S-----D-----EV-----8		Db	1078	GWESDQWQEVIGGEIVKPEQCGVSSGSSLYFSKAGKRLQVSWDLDTSWDFVQFYIQIG	1137

QY 178 ---N-PVSD---LIA-S--GSG---D-S---T-- 192
Db 1138 GESASCNKP--DSREBGLVLOYNNG-GIOWHLLAEMYFSDFKPRFVYLEPAAKTPC 1194
QY 193 AR---I---W---N-L---S-199
Db 1195 TRFRWQPVFSGEDYQWAVDDIIILSEKQKQIIPVINPTLPQNFYKPAFDYPMQMSV 1254
QY 200 ---E--N--ST-S--G-S---T---Q-L---209
Db 1255 WMLANEGWKNTPCAATPSAMIFGKSDGRFAVTRDLTLKPGYVLQFKLNTGCANQFS 1314
QY 210 ---V-LR-H---CIREG---GOD---221
Db 1315 STAPVLLQYSHDAGMSWFLVKEGYPASACKGC--EGNSRELSEPTMYHTG-DFEBWTRI 1371
QY 222 ---VP---S-N---K---D---V---228
Db 1372 TIVIPRSLASKTRFRWIOBSSQKNVPPFGLDGVVISEPCPSYCSGHGDCISGVCFCDL 1431
QY 229 ---T-S---LD---W---N--S---EG--237
Db 1432 GYTAAQCTCVSNPNENMFREFEGLSPLWYKITGAQVGTGCTLNDGKSLYFNGPKR 1491
QY 238 ---TL-LAT---GS---Y--D-G---247
Db 1492 EARTVPLDTNIRLVQFYIGSKTSGITCIKPRTRNEGLIVQYSDNGILWHLLRELD 1551
QY 248 ---F---A-RIW---T---K--254
Db 1552 MSPLFQIISIDLPODAKTPATAFR-WWQPHQKHAQWALDDVLIGMNDSSQTGFQDKF 1610
QY 255 DG---N---LAS-T-L---G-Q---H---265
Db 1611 DGSIDLOANWYRIQGGQVDDICL-SMDTALIFTENIGKPRYAETWDFHVSASTFLOFEMS 1669
QY 266 KG---PIF---AL--K-W--NKK--G---N---279
Db 1670 MGCSKP-FNSHVSQVLOYSLNKGKHLVTEBQVPTTGCLHYTESIYTSERFQWKR 1728
QY 280 ---F-I--LSAGVDKT-TI--I---W---DA---295
Db 1729 TVYLPLSTISPRFRWIOANYTVGAD-SWAIDNVVLASGCPWMCGRGICDAGRCVCDR 1787
QY 296 ---H---T---CE---299
Db 1788 GFQGPYCVVPLPSILKDDFNGLHPLDLWPEYVYGAERGNLNGETIKSGTSLIFKGEGLR 1847
QY 300 ---AK-Q--QF---P---F---H---SA---309
Db 1848 MLISRLDCTNTMYVQPSLRPIAKTSPERSHSLLOFSISGGITWHLMDPEFYPTTNIL 1907
QY 310 ---P---A--L---D---V---314
Db 1908 FINVPLPYTAQTNATFRMLWQPNYNGKKEIWIWDDFIIDGNVNVNPNVMLDTDFDGP 1967
QY 315 D-W---Q---SN---NT--F--A 323
Db 1968 DNWFYFPGNIGLYCPSYKGAPEDSAMVFSVNEVGEHSITTRDLNVNENTIIOFEIN 2027
QY 324 SCSTD---MC---330
Db 2028 GSTDSSADPVLEFRDFGATWHLPLCYHSSHVSLSCTEHHPSYTYAGTMQGW 2087
QY 331 ---IH---V---333
Db 2088 RREVWHGKLHLCGSRFRWYQGYFAGSQPVMTWIDNVYIGQCEMCMNGQSGCINGTK 2147
QY 334 ---CK---L---GO--DR---P---I-KT---F-Q 346
Db 2148 CIGDPYSGPTCKISTKNPDLKDDPEGQLESDFLLMSGKPSRKCGILSSGNLFFNE 2207

QY 347 -G---H---T-N---EV---352
Db 2208 DGLRMLTRDLDLSHARFVQFFMRLGCGKGVDPDRSQPVLLQYSLNGLSLLQEBFLPS 2267
QY 353 ---N---A-I--K---W---357
Db 2268 NSSNVGRYIALEIPLKARSGSTRLRWQPSENGHFYSPWVIDQILIGNISGNTVLBDDF 2327
QY 358 ---D---P---TG---N--LL---ASCS 368
Db 2328 TTLDSRKWLLHPGCTKMPVCGSTGDALVFIEKASTRYVYSTDVAVNEDSLQIDFAASCS 2387
QY 369 --D---DM---T---L-KI---W---376
Db 2388 VTDSCVAIELEYSDVLGLSMHPLVRDCLPTNVBCSRVHLQRIILVSDTFNKMTRITLPLPP 2447
QY 377 ---S---M---KQ---DN---CV---H---385
Db 2448 YTRSQATRFWRHQAPFDKQOTWAIIDNVYIGDCIDMCSGHGRICIQGNCVCDEQWGLYC 2507
QY 386 -D--L-Q--Q--Q--H---390
Db 2508 DDPETSLPTQLKDNFNRAPSSQNLTVNGKLSVCGAVASGMALHPSGCGSRLLVTVDL 2567
QY 391 ---NKE-I---Y---T---IKW---S-PTG---403
Db 2568 NLTNAEFIQPFMYGCLITPNRNQGVLLLEYSVNGGITWLLMEIFVDQYKSKP-GFVNIL 2626
QY 404 ---P-G-TN--N---ML---A---416
Db 2627 LPPDAKEIATRFWRWQPRHDGLDQNDWAIIDNVLIGSADQRTVMLDTFSSAPVQHERSP 2686
QY 410 -NA---N--L---ML---A---416
Db 2687 ADAGPVGRIAFDMFMDKTSVNEHMLFHDCTVERFCDSPDGVMLCGSHDREVIYVTHD 2746
QY 417 ---S--A---S--F---D---421
Db 2747 LPTGEMIQFKISVCKVSEKIAQNIHQVYSTDFGVSNVYLVPQCLPADPKCSGSVQ 2806
QY 422 -S---T---VR---L--W-D---VD--RG---432
Db 2807 PSVFFPTKGKRITYPLPSLVGNVPRFRFYQKYSMDQMAIDNFYLGPGCLDNCRGGDC 2866
QY 433 ---IC---I-HLL-T-KHQ--E-P---V---445
Db 2867 LREQICTDGYSGPNCLYTHLTKFLKRFDSSEIKPDLWMSLEGGSTCTECGILABDTA 2926
QY 446 -Y--S---V---A-F---S-P---DG---454
Db 2927 LYFGGSTVQAVTQDLDLRGAKFLQYWGRISENNTSCHRPICRKEGVLLDYSTDG 2986
QY 455 ---R-Y--L--A---SG---460
Db 2987 WTLHEMDYQKYSVRHDYILLPEDALTNTTLRWQPFVISINGIVVSGVERAQWALDNI 3046
QY 461 ---S---PD---K---C---VHI-W-N--T---471
Db 3047 LIGGAEINPSQLVDTFDDGTSHEENWSPFNNAVTRTAGFCGNPSFHLWPNKKKDKTHNA 3106
QY 472 ---Q--V---C-LH---Y---L--N-479
Db 3107 LSSRELLIQGYMMQPKIVVGCEATSCGDLHSWMLEYTKDARSQNLVQTCPLPSSNS 3166
QY 480 -G--Q--VL--LN---485
Db 3167 IGCSPPQFHEATYNSVSSWKIRITIQLPDHVSSSATQFRWIKGEBETEKQSWAIDH 3226
QY 486 ---L---G---R---SI 490
Db 3227 IGEAPKLCSGHGYCTTGAICIDCESFOGDCSVFSDHLPYSYIKDNFESARVTEANWET 3286
QY 491 ---C---LY---T---494

Db 3287 QGGVIGSCQQLAPYAHGDSIFYNGCOIRQAATKPLDLTRASKIMFVLQIGSMQSDTSCN 3346
Qy 495 --L--PH-----HLV-----V-----IPL-----V----- 505
Db 3347 SDSLGPDAVDKAVLLQYVNNGITWHVIAQHQPDKFTQAQRVSNVPLEARMKGVLLRW 3406
Qy 506 -----AL--IE--LLV-----LK 514
Db 3407 QPRHNGTGDWDALDHVEVVLVSTRKQNYM 3436

RESULT 10
US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING Igg Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-388-9

Query Match 69.9%; Score 2552.2; DB 3; Length 5405;
Best Local Similarity 9.7%; Pred. No. 1.6e-31;
Matches 408; Conservative 17; Indels 3678; Gaps 359;
Qy 1 M-----S-I-----SS-D-EVN-----FLV-----YR-----Y----- 15
Db 1 MGALWSWILWAGATLLWGLTQEAQVDLK-NTGREBFLTAFLQNYQLAYSKEYPRLLISS 59
Qy 16 -----L-O-----ES-----G-----F-S-H----- 23
Db 60 LSESPASVLSQADNTSKVTVRPGESVMWNISAKAEMIGSKIQAHVVIHSDYALISVQ 119
Qy 24 -----SA-----F-----T 27
Db 120 ALNAKPDATLTLRPQALGTEVFLTPPGTSARNVKEFAVVAGAAGASVTLKGSVT 179
Qy 28 F-----G-----I-----K-----S-----HI- 34
Db 180 FNGKFPAGDVLRTLPQYNNVAQLQSSVDLSGSKVTASSPVLVSGHSCAQKHTCNHV 239

Qy 35 -----SQS-----N-I----- 39
Db 240 EQLLPTSAMGTHVVTPLASQSRDYDLAFVVASQATKLTYNHGIGTSRGLOAGDVVEFEV 299
Qy 40 -----N-----GAL--V--P--P--AAL-----I-S----- 51
Db 300 RPSWPLYLSANVGIQVLLFCTGAIRNEVTYDPLVLPDVAAYCPAVVVKVSPGCEGVAL 359
Qy 52 II-Q-K-----G--LOY--V--E--AEV--S--I-----N----- 67
Db 360 VQAQTKAISGLTIDGHAVGAKLTWEAVPGSEFSYAEVGLGTADMIHTAETTLGLLTFG 419
Qy 68 --E-----D-G-T-L----- 72
Db 420 LAKAIGYATAADCGRTVLSPEPSCGMCQAACQRCQVGVGKAGCAEATAVACRAQGDPH 479
Qy 73 ---FDGR-----P---IE----- 79
Db 480 YTTFDGRRYDMGTCSYTWELCSEDDTLPAFSEAKNEHRRGSRVSYVGLVTVRAYSHS 539
Qy 80 -SLS-----LID-----AV-----M 88
Db 540 VSLTRGEVGFVLVDNQRSLPVSLSERLRYQSGPRAVVELVFLVGLVTVTYDWCQALSL 599
Qy 89 P---D-V-----V-----Q----- 93
Db 600 PARFQOVCGLCGNYNGDPADDELTPDGALAPDAVEFASWKLDDGDLCEGQCNCPA 659
Qy 94 -TR-Q-QAY---R-----D-KLA-----Q----- 106
Db 660 CTFQAQHYEGDRLCGMLTKLDGPFVACHDTLDPFPFLEQCVYDLCVVGGERLSLCRGLS 719
Qy 107 --QA-----A-----A-----AAA----- 114
Db 720 AVAAQACLELGISVGDWRSPANCPLSCPANSRYELCGPACPTSCNGAAAFSCNCRPCEG 779
Qy 115 -----A---A---AAS-----Q-Q-G-S-A-K-----NGSN---T----- 130
Db 780 CVCPLPGFVAGGACVPFASCGCTFQGLQAPGQEWADDELQORRCTCNGATHQVTCEDKQ 839
Qy 131 -----ANGE---EN-----GA--H-----T---I---AN- 143
Db 840 SCFA-GERCQVQNGLLGCYDPDRFGTCQSGSDPHVVSFDGRRFDMGTCTVLLVSGCQNA 898
Qy 144 -----N-H-----T-----DMME----- 150
Db 899 ALPAFRLVENEHRGSGTQVSYTRAVREARGVKAVRREYPGQVLDVLLQYLPFOAAG 958
Qy 151 -----V--D-G--D-----V--E----- 156
Db 959 QVQVFRQGRDANVVRTDFGLTVTDNARVTAQVPSYAEALCGLCGNFGNDPADDLALRG 1018
Qy 157 -----I-----P--P-----N----- 160
Db 1019 GGQAANALAFGNSWQBETRPCCGATEPCPKLDSLVAQOLQSKNECGLADPKGPPREC 1078
Qy 161 --K-----AV---VL-R----- 166
Db 1079 HSKLDPQGAVRDCVYDRCLLPQSGPLCDALATYAAACQAAGATVHPWRSEELCPUSCPP 1138
Qy 167 -----C-----G-----HE----- 169
Db 1139 HSHYACSYCCPLSCDGLPFGCGSGECCVCDDEGFPALSGESCLPLASCGCVHOGTYH 1198
Qy 170 -----S-----E-----V----- 172
Db 1199 PGQTFYPGPCDLSLCHCEGGLVSCSSCGPHEACQPSGSGSLGCVAVGSSSTCOASGDP 1258
Qy 173 -----F-----I---C-----AW-----N- 178
Db 1259 HYTTDGRDFDMGTCTVYVLAQTCTGTRPGLHRAVLEQNVAMNGRVSVTRVITVQVANF 1318
Qy 179 -----PVS-----D-L--LASG---S-G-D---ST-----A-----RI-- 195

428 QY -----D-----V-----429
3536 Db FQCPAHSHYELCGDSCPGSCPSLSAPGCSACREGCVCDAGFVLSGDTCPVPGQCCLH 3595
430 QY -----G-I-----CI-----435
3596 Db DDYYPPLGQTFYPPGPGCDLSCRCRGEVSCPSGPHETCRPSGSLGCVAVGSTCQ 3655
436 QY -----H-----TK-----H-----OE-----443
3656 Db ASGDPHYTFDGRREFMGTCVYLAQTCGRPLHRFAVLQENVANGRGVSVTRVITV 3715
444 QY -----PV-----446
3716 Db QVANFTLRLEQRQKVTGVGDMKLPVVLANGIRASQHSDDVVIEDFGLRVAYDLVY 3775
447 QY -----SV-----A-F-S-----PD-----453
3776 Db VRVTVPNGYYQLMCGLCGNYGDPKDDFKPNGSQAGNANFEGNSMEEVVDPSPCLPPPT 3835
454 QY -----G-----RY-----LAS-G-----S-----PD-----463
3836 Db CPPSEGCIPSEBPPLEKXYKKEFCGLLSPTGLPSLCHKLVDPQGPLKDCIFDLCL 3895
464 QY -----K-C-----V-----HI-W-----N-----T-----471
3896 Db GGNLSILCSNIHAYVSACQAGHVPWRNETFCPMECPQNSHYELCACTCSLGSALS 3955
472 QY -----Q-V-----CLH-----Y-----L-N-----479
3956 Db APLOCPDGAEGCQDGSGLYNGACVPICQCGYHNGVYEPBQTVLIDNCRQOCTCHV 4015
480 QY -----GOV-----LL-----N-----LG-----R-----S-----ICL-Y-T-494
4016 Db GKVVVCOEHSCKPQGVQCPGSGILSCVNDKPCGHVTCRPOETCKEQGGQGVCLPNYEATC 4075
495 QY -----PH-H-----LV-----V-----I-----P-LV-505
4076 Db WLWGDPHYHFDGRKDFQTCNVLATTCGPGVSTOGLTFPTVTKNQNRGNPAVSIVR 4135
506 QY -----AL-----I-----E-----L-L-V-----L-----K 514
4136 Db VVTVAALGTNISHKDEIGKVRVNGVLTALPVSADGRISVAQGASK 4182

RESULT 11
US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: /08/471,119A
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15281 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-2

Query Match 59.4%; Score 2533.4; DB 2; Length 15281;
Best Local Similarity 7.7%; Pred. No. 1e-28;
Matches 461; Conservative 46; Mismatches 4; Indels 5466; Gaps 373;
QY 1 M-----S-----ISSDE-----VNF-----L-----V 12
Db 5261 MKATNELSVRYAAVLHI-SDEPLPIYKIDPEAWINFEGSRLTREALAQVLKENENAESV 5319
QY 13 -----Y-----R-----YL-----Q-----17
Db 5320 AISNIPYSKTVVERHIVRSIQDQDANAPESMDGSDWISAVRTRAQOCHTSLASDLFDIA 5379
QY 18 E-SGP-----S-----HS-----A-F-----T-----F-----28
Db 5380 EDAGFRVSVWARQHSQHGALDAVPHHLKPATEDSRVLKFPDTHOGRPLKSLTNQPLLP 5439
QY 29 -----G-----I-----G-----I-----KS-----32
Db 5440 AQSRRAELLIREGLQTLPPYMIPTSLIDRPLNANGKVDRELRARRAKITQSKPVE 5499
QY 33 -----H-----H-----H-----H-----IS-----35
Db 5500 DIVPPRNSVEATVCKGFTDVLGVGVGITDNFFNLGHSLSMATKLAARLGRQLNTRISVRD 5559
QY 36 -----OSN-----I-----I-----N-GA-----42
Db 5560 VFDQPVVADLAAVIQRSAPHEPIKPADYTGVPVQPSFAQGRMLFQDLNNGATWYLMPLG 5619
QY 43 -----L-----L-----V-----P-----P-----46
Db 5620 IRLHGSRLVDALATAISALEQRHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKD 5679
QY 47 -----A-----ALI-----S-----II-----53
Db 5680 AYLAVLKHEQTLFDLATEPGMRVALIRLGEERHILSIVMHHIISDGSVEVLFDENHRF 5739
QY 54 -----Q-----K-----GL-----QY-----V-----E-----AEV-----64
Db 5740 YSSALRQDPMEQILPFIQYRDFAAWQKTEQVASHQRQLDYVTEHLADSTFAELLTDL 5799
QY 65 -----SI-----N-----N-----L-----F-----D-----74
Db 5800 PRFSILSGRANELPLTIEGLHDKLRAFCEVHQATPFVILLAAALRAAHYELTCAEDATLG 5859
QY 71 T-----L-----L-----L-----F-----D-----74
Db 5860 TPIANRNPENLENNIGFFVNTQCMRIAIBENDONFESLVRVRSTATSAFANQDVPFESIV 5919
QY 75 -----G-----R-p-----I-----E-----S-----80
Db 5920 SSLLPGSRDASRNPLVQVILAVHVSQODLGLKLTLEGLRDEAVDSAISTRFDPVEHFLFEHAD 5979

Qy	81	-LS-	--L--I--	--D--	--AVMP-	-----DV
Db	5980	RLSGSVLYAKELFKLRITIESVVSFLETLRRALDQPLPLPLPTDGVGGEIAASKGLLDV				6039
Qy	92	--V-Q-TR-	--QO-	-----QQ-	-----	97
Db	6040	PRTDYPRDANIVEFOQRVRATPDDAIVKDATSILTYAQDDQSDRIAIIWLSRRHMPET	:			6099
Qy	98	--AY-	--R--KL-	-----A-	-----A-	104
Db	6100	LVGVLAPRSCTIIAMGIMKANLAYLPDLINSPAARLSILSAVDGNKLVLLGSGVTAP				6159
Qy	105	-Q-	-Q-A-A-A-	-----A-	-----	110
Db	6160	EQENPEVAAGIOBILLAGTGLDKTGQSNAEPSATSLEYIFTSGSTGPKGVMEHRSVT				6219
Qy	111	--A-	-A-A-A-A-A-	-----A-	-----A-	118
Db	6220	RLAKPSNVIKLPQGARVAHLANIAPDASIEIATILLNGATLVCLDYHTVLCRTELKEV				6279
Qy	119	--S-	-Q-	-----Q-	-----Q-	121
Db	6280	FERESITVTLMPALLQCVAEIPETLAHLDLLYTGDRVGGHDAMEARSIVKIGMPSGY	:	:	:	6339
Qy	122	G-S-A-	-K-N--GE-	-----	-----	128
Db	6340	GPTENTVIITYEDADEMFVNGVPICKTVNSGAYVMNRNQQLVPSSVVGELVVTGDGL				6399
Qy	129	-----N-TANGE-		-----	-----	134
Db	6400	ARGYTDPSLNKNRPYITVNGESIRAYRTGDRVRYRHDLQIEFFGRMDQOVKIRHRIE				6459
Qy	135	-E-N--GA-	-----	-----	-----	138
Db	6460	PGEVESALLSHNSVODAAVICAPADODSGAEWFAVFAARTEDEDTQEESAENDVOQGWE				6519
Qy	139	-H-T-I--ANN-	-----H-T--D-	-----	-----	147
Db	6520	THPETAAYSEVKDIRQSEVGNDFMGWTSMYDGSIEDKTMHEWLNDTMRMLDAREPGHV				6579
Qy	148	-----M-M--E-V--DG-				154
Db	6580	LEIGTGTGMWFNLAKCPGLQGYVGFSPSKSAQAQFNDAQAQSPALKDGRSIVHVGPATD				6639
Qy	155	-----VE-IP-	-----P-	-----	-----	158
Db	6640	INKAGPIQPRLVINSAQFPPTPEYLFRVVEALVQIPSVVERIVFGDMRTNAINRDFVAS				6699
Qy	159	-----	-----P-	-----	-----	159
Db	6700	RALHTLGKANKRLVRQMIYELEANEBELLTPAFFTSRLTREGIKKHVEILLPKTKMAT				6759
Qy	160	N-K--AVUL-RG-H-	-----E-	-----	-----	169
Db	6760	NELSKYRYAAVLHVGRSREQSTHQVSPNAWIDPAADGLDRQTLLINLLKEHKDAGTVAIG				6819
Qy	170	--SE-V-FI-	-----C-	-----	-----	175
Db	6820	NIPYSKTIVERFVNKSISEDMMEEGONSIDGSAAVAAVRMAAQSCPISLDAMDVKEIAQEA				6879
Qy	176	-----A-W-N-P--SD-	-----	-----	-----	182
Db	6880	GYQVEVSWARQWSONGALDAIFHHFEPPEKGARTLIEFPPTYDEGRNVNTLTNRPLNSIQS				6939
Qy	183	-----LL-	-A-S-	-----	-----	186
Db	6940	RLRGTOIREKLOTLLPYPMIPSRIMVLDPMPNNNGIKDKELVRRIVAIVAPKPSRAATRV				6999
Qy	187	-----G-S--G-	-----	-----	-----	189
Db	7000	APRNEIEAILRDEFEDVLGTEVSVLDNFDDLGGHSLMATKLAARVSRRLDAHISIKDVFD				7059
Qy	190	-----D--S--T--A-RIW--NL-	-----	-----	-----	198

DB	7060	QPVLADLAASIQRESAPHEPI	PORPYTGPABQSFAGRLWFLDQLNLGATWYLMPLAIRI	71119
QY	199	-----	-----	198
DB	7120	RGQLRVAALSAALFALERRHETLRT	TFEESDGVGVQIVGEARNSDLRVHDSVTGDDGEYL	7179
QY	199	-----SE-----	-----	200
DB	7180	EVLRRQTVPDFLSSSESGWRVCLVK	TGEEDHVLISVMHHIIYDQWSVDILRGELGQFYSA	7239
QY	201	-----N-----	-----ST-----	203
DB	7240	ALRGDPLHLANPLUPTQYRDFAAWRE	AKOVESHQOLGYWSKQLVDSTPAELLTDLPRP	7299
QY	204	---S---GS---T---	---Q---L-VLR-H---	213
DB	7300	SILSGRAGSDVDVTEIGSVYGALQSF	CTRSTVTFVLLTVFRIAHFRLTAVDATTGTPI	7359
QY	214	-----C-I-----R-----	EG-----	218
DB	7360	ANRNRPELTGVCFVNTQCMRISADDDN	FEGLVRQVRNVATAAYANODVPERIVSAL	7419
QY	219	--G-----	--Q-----D-----VP-----	223
DB	7420	VPGSRNTRNPLVOLMPAVOSVEDYQVR	LEGLESVMMPGEASTRDMFHLVPGQKLT	7479
QY	224	---S---N---K---D---V---	---TS 230	
DB	7480	GSVLYSDLLFEQGTIQNFVDI	FOECLRSVLDQPLTIPSVLPFSNAISNLESLLLEMP	7539
QY	231	-----	LD-----W-----	233
DB	7540	DYPRDRTWDLFREQAACPD	SIANKSSQLTYAQLEQSDRVAAMLHERHMPAESLVG	7599
QY	234	-----N-----	-----S-EG-----	237
DB	7600	VLSPRCETIIAVGIMKANLAYLPLDV	YAPDARLAAILDVTEGERLLLLGAGVPQPGIQ	7659
QY	238	---T---L---LA---	---TG---	243
DB	7660	IPRLSTAYIAEALSHATTVDVTSI	POPSATSLAVITFTSGSTCKPKGVMEHGRGIVRLVR	7719
QY	244	-----S-----	Y-----	245
DB	7720	DTNVNVPESGALPVSHFSLNLA	WDAATWEIYTAVLNGTVVCIDRDTMLDIAALNSTPR	7779
QY	246	-----	D-----	246
DB	7780	KENVRAAFFTPAFKQCLAETPEL	VANLEILHTAGDRDPGDANLAGKTAKGGIFNVLGH	7839
QY	247	-----G---F-----	AR 250	
DB	7840	TENTAYSTFPVVGEEFTFNGV	PGRGISNSHAYIIDRHQKLPAGVMGELIITGDGVAR	7899
QY	251	-----I---W---T---K---	---DG---	256
DB	7900	GYTDSALNKDRFVVYIDINGKSTWS	YRTGDKARVPRDQGLEFFGRMDQVMKIRGVRIEPG	7959
QY	257	-----	-----	256
DB	7960	EVELTLLDHKSVLAAATVVVRRP	PNPGDPEMTAFITIDAEDDVQTHKAIYKHLQILPAYMI	8019
QY	257	-----N---LA---ST---	-----	261
DB	8020	PSHLVILDQMPVTNGVKDRKDLA	LRAQTVQKERSTAARVPPRDEVEALCEEYSNLLLEV	8079
QY	262	-----LG-----	Q-----HK-----	266
DB	8080	EVGITDGFDDLGHSHLLATKLAARLS	QNLNTRVSKVDFOPIADLADIIRRGSHRHPD	8139
QY	267	-----GPI---FA---LKW---	N---KKG-----N-----	279

Db 8140 IPATPYTGPVEQSPAQGL-WFLEQLNGASWYLMPEAIRMRGPLOTQKALAVNALVHR 8198
Qy 280 -----F-----I-LS-AGVD-----K--TT----- 290
Db 8199 HEALRTTFEDHDGQVQVIOFKSSQDLRIIDLSA-VDDTAYLAALKREQTTFADLTSEP 8257
Qy 291 -----II--W----- 293
Db 8258 GWRVSLRLGDDDDYILSVMHHIISDGWTDVLRQELGQFYSAAIRGQEPQLSAKSLPIQ 8317
Qy 294 --D----- 294
Db 8318 YRDFAVQROENQIKEAQKQYWSQQLADSTPCFELTDLPRPSILSGEADAVPMWIDGT 8377
Qy 295 -----AH--TG-----E----- 299
Db 8378 VYQLLTDRCRTHQVTSFVLLAAPTAAHRTAHYRLTGTLDATVGTPIANRNRPELEGLIGFFVN 8437
Qy 300 -----A--K--QO-----F--PF----- 306
Db 8438 TQCRMAISETETESILVQOVRLLTTTEAFANQDVPEQIVSTLLPGSRDTSRNPVQVMF 8497
Qy 307 -----H-----S----- 308
Db 8498 ALQSQDGLRIQLEGMTDEALETPLSTRDLLEVHLFOEVGKLSGSLIYSTDLFEVETIRG 8557
Qy 309 -----A-PAL-----D--V-- 314
Db 8558 IVDVFLBILRGLEQPKORLMAMPITDITKLDRDQGLLTAKPAYPRESSVIDLFRQOVA 8617
Qy 315 --D--W-----OSN-----N--T--FA--SC----- 325
Db 8618 AAPDAIAVWSSSTLTIVADLDGOSNKLAWLHCORNMAPETLVAVFAPRSLTIVAFGLVL 8677
Qy 326 ----- 325
Db 8678 KANLAYLPLDVNAPARIEALSAPGKLVLVQAHPGLTMADETQLVQIDEALASS 8737
Qy 326 -----ST-----D-----M-- 329
Db 8738 SGDEQIHASGPTATSLAYVMFTSGTGKPKGVIMIDHRSIIRLVKNSDVVATLPTPVMA 8797
Qy 330 -----C-----I-----HV--C 334
Db 8798 NVSNLAFDISVQEIYTTALLNGTTLVCLDYTLTLDKILYNVVFVEAQVNAAMFTPVLLKQC 8857
Qy 335 -----KL-----G--QD----- 339
Db 8858 LGNMPAIIISRLSVLFNVGDRDLDAHDAVAASGLIQDAYVYNAVYGTENGQMSTMYKVDVNEP 8917
Qy 340 -----R-----P----- 341
Db 8918 FVNGVPIGRSITNSGAYMVGNOQLVSPGVMEIVVTGDLGARGYTDSDALDEDRFVHVTI 8977
Qy 342 -----IK--T-----P--O-----GH----- 348
Db 8978 DGRNIIKAYRTGDRVRYRPKDFEIEFGRMDQVKIRGHRIEPAEVEHALLGHDLVHDA 9037
Qy 349 -----TN-----E-V--N-----A-IR----- 356
Db 9038 VVLRKANQPEMTAFITTSQDEETIEQESNKQVQGWGEHFDVSRVADIKDLDTSTFGHD 9097
Qy 357 -----W-----D--P-----TG--NL----- 363
Db 9098 FLGWTSMYGDVDPVNEKWEKLETTASLLDNRPFGHILEBIGAGTGMLNSLNGVDGLQK 9157
Qy 364 -----LA----- 365
Db 9158 YVGLDPAPSAAI FVNEAVKSLPSLAGKARVLVGTALDIGSLDKNETQPELVVINSVAQYF 9217
Qy 366 --SC-----S-----D----- 369
Db 9218 PTSEYLLIKVYKAVVEPSVKRVFGDIRSQALNRDFLAARAVRALGDNASKEQIREKIAE 9277

Qy 370 -----D-----M--T-----L-----K-- 374
Db 9278 LEESEBELVDPAFFVLSRQSPNIIKHVEVLPKMKATNELSSYRYAAVLHISHNEEQ 9337
Qy 375 -----I--W-----S-----M-----K-----QD-- 381
Db 9338 LIQIDPTAWDFPAATQKDSQGLRNLLQOGRDDVMIAVGNIPYSKTIIVERHINSLDQDH 9397
Qy 382 ----- 381
Db 9398 VNSLDGTSWISDARSAAICTSPDAPALTQAKEGFRVELSWARQSRONGALDAVFHRL 9457
Qy 382 -----NC-----VH----- 385
Db 9458 ATDANCERSRVLVHFPTDHQGRQLRTLNRPLQARSRRIESQVFEALOTALPAYMIPSR 9517
Qy 386 ----- 385
Db 9518 IIVLPOMPTNANGKVRKQLARRAQVVAKRKAVSARVAPRNDTEIIVLCBEYADILGTEVG 9577
Qy 386 -----DL----- 387
Db 9578 ITDNFDMGSHLMATKLAARLSRRLDTRVTKEVPDKVLADLAASIEQGSTPHLPAS 9637
Qy 388 ----- 387
Db 9638 SVYSGPVEQSYAQRLWFLDQPNLNATWYHMSLAMELLGLNMDALDVALRLEQRHETL 9697
Qy 388 -----Q--Q--H-----N--KE----- 393
Db 9698 RTTFEAKQDIGQVQVHEAGMKRLKVLDSLDKKEKHEHMAVLENEQMPFTLASBPGMKHGL 9757
Qy 394 -----I-----YT-----I----- 397
Db 9758 ARLGPTEYILSLVHHMFSDGWSVDILRQELGQFYSAALGRDPLSQVKPLPIQYRDFAA 9817
Qy 398 --K-----W-----S-----P--T--CP----- 404
Db 9818 WQKEAAQVAEHERQALAYENQLADSTPGELLTDPRPQFLSGKAGVIVPTIEGVPYKLL 9877
Qy 405 -----GT--N--N--P-----N----- 410
Db 9878 KFSKEROVTILFSVLLTAPRATHPRLTGAEDATIGTPIANRNPPELHIIIGFFVNTQCMRL 9937
Qy 411 -----A--N-----LM----- 414
Db 9938 LLDTGSTFESLVQHVRSVATDAYSNQDIPFERIVSALLPGSRDASRPLIQMFALHSOP 9997
Qy 415 -----L--A-----S--A-----S--F-- 420
Db 9998 DLGNITLEGLEHERLPTSVATFDMFHLFQBPNKLSGILFADELFPQETINSVVTVFQ 10057
Qy 421 -----D--ST-----W-----V--R----- 425
Db 10058 EILRGLDQPOVSISTMPLTDGLIDLEKLGLEIESNPRDYVVVDVFRQVAAFNAP 10117
Qy 426 -----L----- 427
Db 10118 AVVDSETSMYSYTLSDQKSEQIAAWLHAQGLRPESLICVMAPRSFETIVSLFGILKAGAY 10177
Qy 428 --DV-----DR----- 431
Db 10178 LPLDVNSPAARTQPIILSEVEGKRLVLLSGIDMQSDRMDVETARIQDILITNTKVERSDP 10237
Qy 432 -----G--ICI-----H----- 436
Db 10298 ASIWEIFTAILNGGALICIDYFTLLDSQALRTTTFEKARVNATLAFAPALLKECLNHPTLF 10357

QY	437	-----	436
Db	10358	EDUKVLYIGGDRLDATAAKIQAOLVKGTVYNAYPTEVMSTIYRLTGDGSEYANGVPIG	10417
QY	437	-----	438
Db	10418	NAVSSSGAYIMDQKQRLVPPGVMGELVSVGGDLARGYNTSLNADRFVDIVINDOKARAY	10477
QY	439	-----T-----K---HQ-EP-----	444
Db	10478	RTGDRTRYRPKDGSIIEFFGMRDQVKIRGHRVPEAEVEQAMLGKAIHDAAVVQAVDQG	10537
QY	445	-----V-----Y-----S--	447
Db	10538	ETEMIGFVSNASDRFSEGEEITNOQWEDHFESTAYAGIEAIDQATLGRDFTSMTY	10597
QY	448	-----V-----A-----P-----S-----P--	452
Db	10598	NGNLIDKAEMEELDDTMQSLLDKEDARPCAEIGTGMVFLPNLKNLGLSESYGIEPSR	10657
QY	453	-----D-----G-----R-YL---A	458
Db	10658	SAALFYDKAAQDFPGLOGKTOILVGTAEIDIKLVDFHPDVVINSVAQYPPFSRYLVQIA	10717
QY	459	-----G--S-----F-----DK-----	464
Db	10718	SELIHMTSVKTIFFGDMRSWATNRDFLVSALYTLGDKATKDIQREVARLEENEDLLV	10777
QY	465	-----C-----VH-I	468
Db	10778	DPAFTSLTSQWPGKYKHVEIILPKRMRTSNELSSYAAVLHICRDGGRNRYGRRVHSV	10837
QY	469	-----W-----NT-----	471
Db	10838	EENAWIDFASSGMDRHVQLVOMLDERDRAKTVAIGNIPHSTINERHTTSLDTEGEGIAQ	10897
QY	472	-----Q-----V--CL-----HY-477	
Db	10898	DSLDSAWQSATKAMAARCPCLSVTELVEIGQAAGFRVSESWARQSOHGALDVVFHILE	10957
QY	478	-----L-N-----G-----Q-----V	482
Db	10958	DDRVGVLINFTDFERLPFPGTGLTSRPLQRIQNRFRFSQIREQLQTLPPYMPVSRIV	11017
QY	483	L-----LN-----LGR-----S-----I--C-----	491
Db	11018	LERMLNANSKVDRKELARKARTLOTIKFSATRVAPRNDIEAVLDCDFQAVLGVTVGVMD	11077
QY	492	-----L---Y-T-----L-----PH-----	497
Db	11078	NPFELGCHSLMATKLAARLSRLDTRVSVKDIQNPQILQDLADVQVQSGSAPEAIPSTPY	11137
QY	498	-----H-----L-V-----V-----	501
Db	11138	SGPVEQSFSGRLWFLDQLNLNASWYHMLASRLRGLRIEALQSALATIEARHESLRTT	11197
QY	502	-----IP-LV-A-----L-I-----B--LL-VLK	514
Db	11198	FEEQDGPVQIVRAARANKQLRIIDVSGTEDAYLAALK	11234

RESULT 12

[illegible]

Db 901 SLLGSSPSDAADLSSVCTGVGNLSQDPYGTOSCHPDTHLAEILMKTLNRLGFTDQAF 960
Qy 133 GE-E-N-----G-----AH-----T-----IAN---N-----144
Db 961 GELEKNDKFLGTSSSENSOPAHHELLCSLQKQLLAFCHINNISENSSVALLKHQLQ 1020
Qy 145 ---H-TD-----I-----MM-----149
Db 1021 LLLPHATDIYSRANLLKESPMWNGEKLVDVIYSAAGSMCLQIYNSLLLLPVSVARP 1080
Qy 150 -----E-VD--G-----D 154
Db 1081 LLSVLLDLLPLDCLNLLPAADLLEDEQLQWPLHGPELIDPAGLPLOPAQSWWLVLD 1140
Qy 155 VE-----I-----P-----159
Db 1141 LERTIALIGRCLGMLQSPVSPPEQDTAYWMKTPLFSDGVEMDTPQLDKMCSLLEVA 1200
Qy 160 ---N-----K-----AV-----163
Db 1201 LSGNEQKPFYKLRPEIAVYVDLALGCSKEPARSLMISMODYAVSKWDSATLSNESLL 1260
Qy 164 ---VL-----R--G--HESEVF-----I-----C-----175
Db 1261 DTVSRFVLAALLKHTNLLSOACGESRYQPGKHLSEVYVCYKVSRLACKNLELIQTRS 1320
Qy 176 -A--W-----177
Db 1321 SSRDRWISNQDSADVDPOEHSFTRTIDEEAEEMEQAEARDREBCHPEPEDEBERHEVM 1380
Qy 178 -----N-----P-----VS-----D-----182
Db 1381 TAGKIFOCFLSAREVARSRRDRMNSGAGGARADDPPOSQOERRVSTDLPSGQVYTA 1440
Qy 183 -----LL-----AS-----186
Db 1441 ACNSVIHRCALLILGVSPVIDELQKREEQLOQPSTASGEGGLMTRSESLTAESRLVH 1500
Qy 187 -----G--SG-----D-----S-T-A 193
Db 1501 TSPNYRLIKRSBSDLSPQESDEBEGYALSGRQNVLDLAAASHRKGPMHQSLESLSWSA 1560
Qy 194 RI-----W--N-----L--S-----EN-----STS- 204
Db 1561 RLKHSRDWLCNSSYSFESDFLTKSLGVHTLIENVSVFSGDVGNAPGFKPEESMSTSP 1620
Qy 205 -----I-----GS-----T-----Q 208
Db 1621 QASTIAEQOQLRAELRLALHQLVLLSGWEEKGSISLAGSRLLSGFQSQSTLLTSVRLQ 1680
Qy 209 -----L-----V-----L 211
Db 1681 FLAGCFGLTGVHTGAKGESRLHYQDGIARAARNIQUIEIVAVHKIYQOLSATLAL 1740
Qy 212 -----R-----H-----C-----214
Db 1741 QANKHHIEAQRLLLVTFVFSLVHYQVDVSVLSAISTGLNLVLSQLCGTDTMLGQPLQLP 1800
Qy 215 -----I--R-----216
Db 1801 KTGVSQSLTALKVASTELLQILAITTGYADKLSPKVQSVLLDLCQKLLSQTGVLLH 1860
Qy 217 ---EG-----G-----Q-----D-----V-----222
Db 1861 MASFGEQEGDEBEEKKVDSSGETEKDPFRAALRKQHAELHGLDFLPLFRVRSKAI 1920
Qy 223 -----P-----S--N-----X-D-----V 228
Db 1921 QSKWASPKWTEVILNTASQKCSGIFLVGNLRTLLALHVLBAVLFACEGSGVEDDQMAQI 1980
Qy 229 T-----SL--D--W-----N-----S-----235

Db 1981 VERFLSLLSDCMWETPIAQAKHAIQIKEQEIKLOKQGELEEDENLPIQEVSPDEKA 2040
Qy 236 ---E-G--T-----L-LA-TG-----243
Db 2041 QCCLVENGQILTHSGSGKGYGLASTGVTGCVQWKFYIVKENRGNEGTCTGVSRWPVHDF 2100
Qy 244 ---S--Y-----D-----G-----247
Db 2101 NHRTTSDMMLRAYSGNLYHNGEQTLTSSFTQGDFTICVLDMEARTISFGKNGEBPKLA 2160
Qy 248 -----F-----A-----249
Db 2161 FEDVDAELYPVCMFYSSNPGEKVKICDMQMRGTPRDLPLGDPICSPVAALAEATQLV 2220
Qy 250 RI-----WT-----KD-G-----N-----257
Db 2221 RILHRDRTWYCIKKNOMERLHKIKICESGOKLKKRSVQSREBENREEMKESEEEK 2280
Qy 258 ---LA--S--TL-----G-----Q--H--K- 266
Db 2281 GKTRHGLADLSELQURTLCIEVWPVLAIVGGVDAGLRVGGRCVHKQTRHATLLGVVKE 2340
Qy 267 G-----P-FALKW-----N-----275
Db 2341 GSTSAKVQWDEAEITISFPTF---WSPSDTPLYNLEPCBPDPDVARFRGLTASVLLDLT 2397
Qy 276 ---KK-----G-----278
Db 2398 YLTGVHEDMGKOSTRHEKHRESEBEGDVEQKPESESALDMRTGLTSDDKSQSTTSS 2457
Qy 279 ---N--F-----I-----281
Db 2458 KSENEIASLSLPTPLPSVESQHQITEGKRKNHHSKNHDVAQSEIRAVQLSVLYLGAMK 2517
Qy 282 -LSA-----284
Db 2518 SLSALLGCSKYAELLIPKVLAEHNSDCASSPVVHEDVEMRAALQFLMRHVMKRAVMR 2577
Qy 285 ---G-----V--D--K-----288
Db 2578 SPIKRALGLADLERAQAMITYKLWVHGLLEDQFGKIKQBIDQQAESDPAQAQTPTVTS 2637
Qy 289 ---TT-----290
Db 2638 PRASSTTSFMSSSLEDTTATTPTVDTETVPASESPCVMPLSLLRQMFSSYPTTIVLPTR 2697
Qy 291 -----290
Db 2698 RAQTPPTISLPTSPSDEVGRROSLTSPDSQSARPANRTALSDPSSRLSTSPPPPAIAVPL 2757
Qy 291 ---I-----I--W-----293
Db 2758 LEMGFSRLQIAKAMEATGARGEADAQNITVLAMWMIHPEHGEDDEEPOSGSTADSRPGA 2817
Qy 294 ---DA-----HT-----GE--A--300
Db 2818 VLGSGKSNDCPYLQSPGDI PSADAAMEEGFSESPDNLHTENAASGGSPSARGSAVT 2877
Qy 301 ---K-----QO-----303
Db 2878 RRRHKPDLAARTLLARAAGLYRSVQAHNRQSRREGISLQDDPGALYDFNLDELEIDLDD 2937
Qy 304 ---F-----P-----F-----306
Db 2938 AMEAMPQDLTSDNDILGMWIPVLDWPTWHVCESEDREVVVCELCECSVVSFNQMKR 2997
Qy 307 -H--SA-----P-----310
Db 2998 NHPGCRSANROGRNSGVVDGWFGEGCGSGNPYYLLCGTCKEYKAMTKSKTSSESR 3057
Qy 311 ---A--L--D-V--DW-----Q--317
Db 3058 YKQAPDLIGKQDSVVEEDWMLDVEDEKLAGEBEFELLAGFLGLNDRRIIVPEPVQFPD 3117

Qy	318	-S-	N	-N-	-N-	-N-	320
Db	3118	SDPLGASVAMVTATNSMEETLMQIGCHGSVEKSSGRITLGEQAALANPHDRWALRRV					3177
Qy	321	T-FA-	-SCS-	TD-	MCI-		331
Db	3178	TAAAOVLARTVMWRALSLLSVSGSSCSLAAGLESGLDTRIVLRMLCIAAAGRAGLST					3237
Qy	332	-H-	V-CXL-	G-	O-		338
Db	3238	SPSAMASTSRSGHGSHKANKPI-S-LAYLSTAVGCCLASNAPSAAKLVLQCTONLSIAA					3296
Qy	339	-D-R-P-I-K-T-P-O-					346
Db	3297	TGVNLTTVDSDIORKFPSFLRGIAENKLVTSPNFVVTOALVALLADKGAKLRPNYDKS					3356
Qy	347	-G-	H-	T-N-			350
Db	3357	EVEKKGLELANALAACCLSRSLSSHQRWAQQOVLRTLAARDNQTLTQTADMGDDL					3416
Qy	351	-E-V-N-A--IK-W-					357
Db	3417	RKCSFIKEAHQNRMTCVMCNKKGLLATSGNDGTIRVMNVTKKYQSLOQTCVFNRLEGD					3476
Qy	358	-DPT-	G-N-I-LA-				365
Db	3477	AESLGPSPDPFSPVSWISGKYLAGEAKEMVINQVNGGKLVDIOPHHVSALAPPEE					3536
Qy	366	--S-C-S-D-					369
Db	3537	GPATANGSEPELLVCGRMDGSLGLEVDVSTVHRRELEHCYRKDVSVTCAWFSEDPRP					3596
Qy	370	-D-M-T-L-K-					374
Db	3597	FVGYPDGKLLGTKEPLEKGGIVLIDAHKDTLISMKWDP TGHI LMTCAKEDSVKLWGI					3656
Qy	375	I-W-S-M-KOD-					381
Db	3657	SGCWCLHSLCHPSIVNGIACWRLPGKSKQLLLMATGOSGLVGVWVRIPQDFTQTNVTS					3716
Qy	382	-N-CV-H-					385
Db	3717	AEGWDOESCQDGYRKSSGAKCVYQLRGHTPVRTVPASPDSGLALVSGGLGMNIWSL					3776
Qy	386	-D-Q-H-N-KEY-TIKW--S-					400
Db	3777	RDGSVLQTVVIGSAIQTTWMIPEVGVAAACSRSKOVLLVNCTAEAAAAHNHLATCRTL					3836
Qy	401	-P-T-G-P-					406
Db	3837	KQGVLGLNAPCMRAPLEBLPMWLQEYAYEKPHVVCGDQLVHSPMQCLASLAVGLHL					3896
Qy	407	-N-P-NAN-L-MLA-SA-SF-					420
Db	3897	DQLLCNPVPVPHQCNCPLDPASNPNNEWAMECFSTTIKAEEALTGAQFPESFTVPDL					3956
Qy	421	-DSTURL-W-D-V-D-R-G-					432
Db	3957	PVPEDELVFLMDSKWTNGMDEQIMSWATSRPEDWHLGKCDVYLWGAGRGHQLAEAGRN					4016
Qy	433	-IC-I-H-					436
Db	4017	VMVPAAPSPSQAOQVICGNQCTFVIQANGTVIACGEGSVGRLGQGNSSDLLHLVTVISAL					4076
Qy	437	-T-LT-K-H-Q-EP-					444
Db	4077	QGFVVTQLVTS CGSDGHSMALTSESGEVFSNGDGYGKLGHSNDRQRP RPQIEALQGEEV					4136
Qy	445	V-Y-S-VA-FS-PD-GR-Y-					456
Db	4137	VOMSCGPKHS AVVTS D G K L F T F G N G D Y C R L G L G N T S N K K L P E R V T A L E G Y Q I G O V A C G L N					4196

RESULT 13

US-08-460-751-2
Sequence 2, Application US/08460751
Patent No. 5891628
GENERAL INFORMATION:
APPLICANT: Readers, Stephen
APPLICANT: Schneider, Michael
APPLICANT: Glucksmann, Sandra
TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,751
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/413,580
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 7638-005
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4303 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-460-751-2

Query Match 69.3%; Score 2527.4; DB 2; Length 4303;
 Best Local Similarity 10.0%; Pred. No. 1.5e-31;
 Matches 410; Conservative 82; Mismatches 14; Indels 3576; Gaps 355;

Qy	1	MSIS	-----S-----D-----D-----E-----V-----F-----10
Db	72	LDVSHNLLRALDVCLLANLSALAEIDISNNKISTLEEGIFANLNFLENSEINLSGNPFECDC	131
Qy	11	LVY	-----R-----Y-----15
Db	132	GLAWLPQWAEQQVVRVQPEAATCAGPGSLAGQPLLGLPLDSDGCGEYVACLPLDNSSGT	191
Qy	16	-----I-----O-----E-----16	
Db	192	VAAVFSAAHEGLQPEACAFCTGCGLAALSEQGWCICGAAQPSASFACLSLCSGP	251
Qy	21	-----F-----S-----H-----S-----AF-----T-----FG-----29	
Db	252	PAPAPTCTGPTLLOHVFPASPAGNATLVGPHGLASGQLAAFAHIAAPLPVTDTRWDFDGS	311
Qy	30	--I-----KSH-----I-----S-----QSN-----38	
Db	312	AEVDAAGPAASHRYLPGRYHVTAVLALGAGSALLGTDVQVEAAPAALELVCPSSVQSD	371
Qy	39	-----I-----N-----G-----AL-----V-----P-----P-----46	
Db	372	SLDLSIQNRGGGLEAAYSIVALGEEPARAVHPLCFSDTEIFFGNGHCYELVVEKAWLQ	431
Qy	47	-----AA-----LIS-----I-----I-----Q-----K-----G-----L-----57	
Db	432	AQEQCAWAGALAMWDSPAVQFLVSRVTRSLDVLWIGFSTVQGVGVPAPQGEASLES	491
Qy	58	-----Q-----YV-----E-----AE-----V-----S-----65	
Db	492	CQNWLPGEHPATAEHCVRILGPTGWCNTDLCSAHPSYVCELPQGGPVQDAENLLVGAPSG	551
Qy	66	-----IN-----EDG-----TL-----F-----73	
Db	552	DLOGPLPLAQDGLSAPHEFVEMVPPGLRLSREAFLLTAETGTOELRRPAQLRLQVYR	611
Qy	74	-----D-----GR-----P-----I-----78	
Db	612	LLTAGTTPENGSEPSRPNRTQLAPACMPGGRWCPCGANICLPDASCHPQACANGCTS	671
Qy	79	-----E-----S-----LSLI-----DAV-----M-----P-----DVV-----92	
Db	672	GPGLGAPYALWREFLFSVPAGPPAQYS-VTLHGQD-VLMLPGDLVLQHDAGFGALLHC	729
Qy	93	-----Q-----T-----R-----Q-----Q-----97	
Db	730	SPAPGHGPPAPYLSANASSWLPHLPAQLEGTWGPCACALRLLAQREQLTVLLGLRPNPG	789
Qy	98	-----A-----Y-----RD-----KL-----A-----Q-----105	
Db	790	LRLPGRVEVRAEVGVSRHNLSCSFDVSPVAGRLVIYPAPRDGRLYVPTNGSALVLQV	849
Qy	106	-----QQ-----A-----A-----A-----A-----A-----112	
Db	850	DSGANATATARPWPGSLSARFENVCPALVATFVPACFWETNDTLFSVVALPMLSEGEHV	909

Qy	113	-----A-----A-----A-----A-----A-----A-----118	
Db	910	DVVVENSASRANLSLRTABEPCIGLRATPSPARVLQGVLYRSPVVEAGSDMVRWTI	969
Qy	119	-----S-----Q-----Q-----Q-----Q-----Q-----124	
Db	970	NDKQSLTFQNVFNVYQSAAVFKLSLTASNHNVSNTVNVNVTVERMNRQGLQVSTVPA	1029
Qy	125	-----KN-----GE-----N-----N-----TA-----131	
Db	1030	VLSPNATLALTAGVLVDSAVEAFLWTFDGEQALHQFOPPPYNESFPVDPSPVAQLVEH	1089
Qy	132	N-----GE-----EN-----G-----A-----H-----139	
Db	1090	NVTHTYAAPGEYLLTVLASNAFENLTQQVPVSVRASLPSVAVGVSDGLVAGRPVTFYPH	1149
Qy	140	-----T-----I-----ANNHT-----146	
Db	1150	PLPSPGGVLYTWDGPGSPVLTSQSPAANHTYASRGTYHVRLEVNNTVSGAAAQADVRF	1209
Qy	147	-----DM-----M-----E-----V-----D-----GD-----154	
Db	1210	EELRGLSVDMSLAVEQCAPVWSAAVQTGDNITWTDFMGDGTVLSGPEATVEHVYLAQN	1269
Qy	155	-----VE-----P-----P-----NKA-----162	
Db	1270	CTVTGAGSPAGHARSLLHVLVFLVLEVRPEAACIPTQDARLTAYVTGNPAHYLFDWT	1329
Qy	163	-----V-----V-----L-----R-----G-----H-----EDEV-----172	
Db	1330	FGDSSNTTVRCGPTVTHNTRSGTFLALVLSRRVRAHYFTSICVEPEVGNVTLOPER	1389
Qy	173	-----FI-----CAM-----NP-----VS-----D-----LL-----AS-----186	
Db	1390	QFVLGDEAWLVACAMPFPFYRYTWDFTGEAAPTREARGPEVTFIYRDPGSYLVTVTASN	1449
Qy	187	-----GS-----190	
Db	1450	NISAANDSALVEQEPVLTSTIKVNGSLGLELOQPYLPSAVGRGRPASVYLDLGDGWL	1509
Qy	191	-----S-----TARI-----WN-----L-----SE-----N-----S-----202	
Db	1510	GPBVTAYNSTGDTVRVAGMNEVSSEAWLNVTKRRVRLGVVNASRTVPLNGSVSPS	1569
Qy	203	TS-----GS-----TOL-----VL-----RHC-----I-----RE-----G-----QD-----221	
Db	1570	TSLEAGSDVRSWVLCDR-CTPIPGGPTISYTFRSVGTENIIIVTAENVEGSAQDSIFVYV	1628
Qy	222	-----V-----PSN-----KD-----V-----T-----SL-----231	
Db	1629	LQIEGLQVGGGRYPPTNHTVQLQAVVRDGTNVSVSWTAWDRDRGALLAGSGKFSLTVL	1688
Qy	232	-----D-----W-----N-----234	
Db	1689	EAGTYHVQLRATNMLGSAWADCTMDPVEPVGMLVWAASPNPAAVNTSVTLAELAGSGV	1748
Qy	235	-----S-----EG-----T-----L-----LAT-----GS-----YD-----246	
Db	1749	VYTWLSLEGLSWETSPTTTHSPPTPLGLHLVTMTAGNPLGSANATVEVDVQVPVSGLSIR	1808
Qy	247	-----G-----F-----A-----R-----I-----W-----TK-----D-----G-----NL-----258	
Db	1809	ASEPGGSFVAAGSVFPWQQLATGNTNVSWCNAVPGSSSKRGPHVTWVPDPAGTFSIRLN	1867
Qy	259	AS-----T-----L-----L-----QO-----H-----K-----G-----267	
Db	1868	ASNAVSWVSATYNTLTAEPIVGLVWASSKVVAQGLVHFQILLAAAGSAVTTFELQVGGAN	1927
Qy	268	P-----I-----F-----AL-----K-----W-----N-----KKG-----278	
Db	1928	PEVLPGPFHSHPFRVGDHVSVRGKNHVSQAQVRIIVLVEAVSGLVQVNCCEPGIATG	1987

QY 279 --NF--ILS--AG-- 285
Db 1988 TERNFTARVQGRVAVYAFSLQKQDGLVILSGRDVTVYPAAGLLEIQVRAFNALG 2047
QY 286 --V-D--K--T--T--II--WD-- 294
Db 2048 SENRTLVLEQDAVQVVALQSGCFNRSQAFAATSPSPRRVAYHWDGDSGPGQDTDE 2107
QY 295 --A-HT--G--E--A--KO--Q--FP-- 305
Db 2108 PRAEHSYLRPGDYRVQUNASNLVSFFVAQATVTVQVLACREPEVDVVLPLOVLMRRSQRN 2167
QY 306 F--H--SA--P--AL--DV-- 314
Db 2168 YLEAHVDLRDCVYQTEYRWEVYRTASQCRPGPARVALPGVDSPRLVLPRLALPVGH 2227
QY 315 --D--W--QS--N-- 319
Db 2228 YCFVFWVSGDTPLTQSIQANVTVAPELVPILIEGGSYRVMSTDRDLVLDGSESDPNLE 2287
QY 320 --N--TF-- 322
Db 2288 DGDQPLSFHWACVASTOREAGCALNFGPRGSSVTIIPRERLAAGVEYTFSLTWKAGR 2347
QY 323 --A--SC-- 325
Db 2348 KEEATQTVLIRSGRPIVSLCVCCKAQAVVEVSRSSVYVLEGRCLNCSSGSKRGWAA 2407
QY 326 --S--T--D--M--C--I-- 331
Db 2408 RTEFNKTLVLDTTTTSGAGMRLVLRGVRDGEYTFTLVLRSGSEEGCASIRLSP 2467
QY 332 --H--V--C-- 334
Db 2468 NRPLGSGCRLPFLGAVHALTTKVHFECTGHWDAEDAGAPLVYALLRRCRQGHCEFCV 2527
QY 335 -K--LG-- 337
Db 2528 YKGSLSYGAVLPPGPRHFEVGLVAVVQDLGAVALNRSIAITLPEPNSATGLTW 2587
QY 338 --Q-D--R--P-- 341
Db 2588 LHGLTASVLPGLLRQDPQHVIEYSALVTVLNEYERALDVAEPKHQRAQIRKNIT 2647
QY 342 --I--K-T--FO-- 346
Db 2648 ETLVSLRVHTVDIQIAAALAQCMGFSRELVCRCCLKQTLHLEAMMLILOAETTAGTV 2707
QY 347 --G-H--T-- 349
Db 2708 TPTAIGDSILNITGDLIHLASSDVRAPOPSLGAESPSRMVASQAYNLTSALMRLMRSR 2767
QY 350 --NE-- 351
Db 2768 VLNEEPLTLAGEEIVAGKRSDPSRLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIF 2827
QY 352 -V--N--AI--K--W-- 357
Db 2828 LVDSNPPFPYISNYTVSTKVASMAFQTOGAQIPIERLASERAITVKVPNNSDWAARGH 2887
QY 358 --D--P-TG--N--LLA-- 365
Db 2888 RSSANSANSVVVQASVGAVTLDSSNPAAGLHLQNLVTLDDGHVLSPEPEYLAVALH 2947
QY 366 S--CS--D-- 370
Db 2948 SEPRPNEHNSAGRRIRPESLOQADHRPVTFTISPGSRDPAGSYHLNLSHFRWSALQVS 3007
QY 371 --M-T-L--KIW-S--MK--Q-- 380
Db 3008 VGLYTSLCQYFSEDMVMWRTGGLLPLEETSPROAVCLTRHLTAFGASLFPVPPSHVRFP 3067
QY 381 --D-N--C--V--H-DLQ--QHN--K-EIYTIK 398

Db 3068 EPTADVNYITVMLTCAVCLVTYVMAAILHKLD-QLDASGRAIPFCQGRGRFYEL-VK 3125
QY 399 --W--S-PT--G--P--GT-- 406
Db 3126 TWGRSGGTTAUVHIMLYGVDSRSGHRHLDGRAPHRNLDIPRIATPHSLGSLGWKIRVW 3185
QY 407 -NN--P--N--AN--L--M--L-AS-AS-- 419
Db 3186 HDNKGSLPAWFLQHVIVRDLQTAARSAPFLVNDWLSVETEANGGLVEKVEVLAASDAALLRF 3245
QY 420 --FD--S-T-V-R--L--W--D-- 428
Db 3246 RRLVAELQRFKDHIMLSIWDPRPSRFTRIQATCCVLLICLFLGANAVMYGAVGDS 3305
QY 429 --VD--R--G-- 432
Db 3306 AVSTGHVSRSLSPSLVDTVAVGLVSSVVVVPVYLAIFLFRMSRSKVAGSPSPTPAQOVL 3365
QY 433 -I--C--I--H--TJT-- 439
Db 3366 DIDSCLDSSVLDSFLTFTSGLHAEQAFVGMKSDLFLDDSKSLVCMWPSGEGTILSWPDLLS 3425
QY 440 --K-- 440
Db 3426 DPSIVGSNLRQLARGAQGHGLGPEEDGSLASPYSPAKSFSASDEDLIOQVLAEGVSSPA 3485
QY 441 --H--Q--E-- 443
Db 3486 PTQDTHMETDLLSSISTPGKETTALQRLGELGPPSGLNWEPQAAARLSRTGLVEGL 3545
QY 444 --P--V-Y--SVA-- 449
Db 3546 RKELLPAWCASLAHGLSLLLVAVAVSWGVGASPPGVSVANWLLSSSASFLASFLGWEP 3605
QY 450 --FS--PD-- 453
Db 3606 LKVLLEALYFSLVAKRLHDEDDTLVESPAVTPVSARVPRVRPPHGFALFLAKEARKVK 3665
QY 454 --G-R--Y--LAS-GSPD-KC--V--H--I-- 468
Db 3666 RUHGMRLSLVYMLFLVTLTASYG--DASCHGHAYRLOSAIKQELHSAFIAITRSEEL 3723
QY 469 --W--N-T--QV--C--L-- 475
Db 3724 WPMMAHVLLPYVHGQSSPELGPRLRQVRLQELALYDPDPGPRVHTCSAAGFSTSDYDV 3783
QY 476 --H--Y--L--NG-Q--VL--L 484
Db 3784 GWESPHNGSGTWASAPDLGAWSGSCAVYDSGGYVQELGSLSESRDLRFLQLHNWL 3843
QY 485 -N--L--GR-- 488
Db 3844 DNRSRAVLELTRYSPAVGLHAAVTLRLEFFAAGRALAALSVRPFALRRLSAGLSPLLT 3903
QY 489 SICL--Y--T--L-- 495
Db 3904 SVCLLLFAVFAVAEARTWHREGRWRLVGLGAWARWLLVALTALVRLAQIGAARQW 3963
QY 496 --P--H--H-- 498
Db 3964 TRFVRGRPRFTSFQVAVHVSAAAGLAASLLFLLLVKAQHVRFVRQWSVFGKTLCLRAL 4023
QY 499 --LVV--I-- 502
Db 4024 PELLGVTGLVGLVAYAQALAILLVSSCVDSLWSVAQALLVLCPTGLTSLCPAESWHLS 4083
QY 503 PL--V--AL--I--ELL--V--L 513
Db 4084 PLLCGLWALRWGLALRGAVILRWYHAIIRGELYRPAWEPQDYEMVELFLRRLRLWML 4143
QY 514 -K 514

Db 4144 SK 4145

RESULT 14

US-09-479-467A-2

; Sequence 2, Application US/09479467A

; Patent No. 6723557

; GENERAL INFORMATION:

; APPLICANT: Sternberg, Paul W.

; APPLICANT: Barr, Maureen M.

; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE HOMOLOGS REQUIRED FOR MALE MAT

; TITLE OF INVENTION: BEHAVIOR IN NEMATODES AND ASSAYS BASED THEREON

; FILE REFERENCE: 18021-2901B

; CURRENT APPLICATION NUMBER: US/09/479,467A

; CURRENT FILING DATE: 2000-01-06

; PRIOR FILING DATE: 60/115,127

; PRIOR FILING DATE: 1999-01-06

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 4303

; TYPE: PRT

; ORGANISM: Homo sapiens PKD-1 protein

US-09-479-467A-2

Query Match 69.3%; Score 2527.4; DB 4; Length 4303;

Best Local Similarity 10.0%; Pred. No. 1.5e-31;

Matches 410; Conservative 82; Mismatches 14; Indels 3576; Gaps 355;

Qy 1 MSIS-----S-----D-----E-----EVN-----F-----10

Db 72 LDVSHNLLRALDVCLLANLSALAEIDISNNKISTLEBIEFANLPNLSEINLSGNPFECDC 131

Qy 11 -LVY-----R-----Y-----15

Db 132 GLAWLPQWABEQVRVVPPEAATCAGPGSLAGPILGILLDSGCGEYVACLPDNSSGT 191

Qy 16 -----LQ-----E-----SG-20

Db 192 VAAVSFAAHEGLLPACSAFCFSTQGLAALSEQWCLCGAQPSSASFACLSLCSGP 251

Qy 21 -----F--S-----H--S-----AF-----T--FG--29

Db 252 PAPAPTCRGTLLQHVFPASPGATLVGPHGLASGLAAFAHIAAPLPVTDTRWDFDGS 311

Qy 30 --I-----KSH-----I-----S-QSN-38

Db 312 AEVDAAGFAASHRYVLPGRYHTAVLALGAGSALLGTDVQVEAAPRAALELVCPSSVQSD 371

Qy 39 --I-N--G-----AL-----V--P--P-----46

Db 372 SLDSLQNRGSGLEAAYSIVALGEEPARAVHPLCPSDTEIFPONGHCYRLVVEKAWLQ 431

Qy 47 -----AA-----LIS-----I--IQ-K-----G--L--57

Db 432 AQEQCQAWAGALAWDSPAQRFLVSRVTRSLDVMIGFSTVQGVGVPAPQGEAFSLES 491

Qy 58 -Q-----EDG-----YV-B-----AE--V--S--65

Db 492 CONWLPGEPHATAEHCVRILGPTGWCNTDLCASPHSYVCELQPCGPFVQDAENLLVGAPSG 551

Qy 66 -----IN-----EDG-----TL--F-----73

Db 552 DLQGLTFLAQDGLSAPHEVVEWVPPGLRLSREAFITAEFGTQELRRPAQLRLQVYR 611

Qy 74 -----D-----GR--P--I-----78

Db 612 LLSTAGTPENGSEPSRSPNRTQLAPACMPGGRWCPCGANICLPDASCHPQACANGCTS 671

Qy 79 -----E--S-----LSLI-----DAV-M-P-DVV-----92

Db 672 GPGLPGAPYALWREFLFSVPAGPPAQYS-VTLHGQD-VLMLPGDLVLQHDAGFGALLHC 729

Qy 93 -----Q--T-----R--Q--Q-----97

Db 730 SPAPGHPGPRAPYLSANASSWLPPLPAQLEGWGCPCACALRLLAQREQTLVLLGLRPNPG 789

Qy 98 -----A-----Y--RD-KL-----A--Q--105

Db 790 LRLPGRYEVRAEAVGNGVSRHNLSCSPDVVPVAGLVVYAPAPDGRGLYVPTNGSALVLQV 849

Qy 106 -----QQ-----A--A-----A-----112

Db 850 DSGANATATARWPGGSLSFARFENVCVPCALVATFVPCPFWETNDTLFSVWALPMLSEGHVV 909

Qy 113 -----A--A-----A-----A-----118

Db 910 DVVENSASRANLSLRVTAEEPCGLRATPSPEARVLQGVLYRSPVVEAGSDMFRWTI 969

Qy 119 -----S-----Q-----Q-----S--A 124

Db 970 NDKQSLTFQNVVFNVIYQSAAVFKLSLTASNHVSNVTNVNVTVERMNRMQGLQVSTVPA 1029

Qy 125 -----KN-----GE-----N-----TA-----131

Db 1030 VLSPNATLALTAGVLVDSAVEAFLWTFDGGEOALHQFPFYNESFPVDPSPVAQLVEH 1089

Qy 132 N-----GE-----EN-----G--A-----H 139

Db 1090 NVTHYTAAPGEYLLTVLASNAFENLTQQVPSVRASLPSVAVGVSGVLVAGRPVTFYPH 1149

Qy 140 -----T-----I-----ANNHT-----146

Db 1150 PLPSPGGVLYTWDFDGGSPVLTSQSPAANHTVARSCTYHVRLEVNNTVSGAAAQADVRF 1209

Qy 147 -----DM-M-E-----V-----D-GD-----154

Db 1210 BELRGLSVDMSLAVEQAPVWVSAAVQTGDNITWTFDMDGTVLSGPEATVEHVYLAQN 1269

Qy 155 -----VE-----IP--P-----NKA-----162

Db 1270 CTVTVGAGSPAGHARSLLHVLVFLVLEVRVEPAACIPTQDARLTAYVTGNPAHYLPDWT 1329

Qy 163 -----V-----V--L-R-G-----H-----ESEV-----172

Db 1330 FGDGSSNTTVRGCPVTTHNFTSRGTFPLALSLSRVRAHYFTSICVEPEVGNVTLQPER 1389

Qy 173 -FI-----CAW-----NP-VS-----D--LL--AS-186

Db 1390 QFVQLGDEAWLVACAMPFPYRYTWDFTGTEEAAPTRARGPEVTFIYRDPGSYLVVTASN 1449

Qy 187 -----GS-----GD-----190

Db 1450 NISAANDSALVEQEPVLTYSIKVNGSLGLELQQLPFLSAVGRGPASVYLDLGDGWL 1509

Qy 191 -----S-----TARI--WN-L-SE-----N-----S 202

Db 1510 GPEVTHAYNSTGDTFVRVAGMNEVSERSEAWNLVTKRRVRLVWNASRTVPLNGSVSFS 1569

Qy 203 TS--GS-TQL--VL--RHC-----I--RE-G-----G--QD-----221

Db 1570 TSLEAGSDVRYSKVLCDDR-CTPIPGPTISYTFRSVGTFTNIIIVTAENVEGSAQDSIFVV 1628

Qy 222 -----V-----PSN-----KD--V--T-----SL--231

Db 1629 LQLIEGLQVVGGRYFPNTNHTVQLQAVVRDGTNVSYSWTAWDRDGPALAGSGKFSLTVL 1688

Qy 232 -----D-----W-----N-----234

Db 1689 EAGTHVQLRATNMLGSAWADCTMDPVEPVGLMWAAASPNAAVNTSVTSLAELAGSGV 1748

Qy 235 -----S--EG-----T--L-LAT-----GS-----YD-----246

Db 1749 VYTWSLEEGSWETSEPFTHSPPTPGLHLVMTAGNPLGASANATVEVDVQVPSGLSTR 1808

Qy 247 -----G-F--A-----R-----I--W-----TK--D-G-----NL 258

Db 1809 ASBPGGSFVAAGSVFPWGQATGTNTVSWCWAIVGGSSKRGPHVTWVFPDAGTFSIRLN- 1867
QY 259 AS- - - - -T- - - - -L- - - - -GO- - - - -H- - - - -K- - - - -G- - - - - 267
Db 1868 ASNAVSVSATYNTLTAEPIVGLVWASSKVAPGQLVHFQIILLAGASAVTFLQVGGAN 1927
QY 268 P-I- - - - -F- - - - -AL- - - - -K- - - - -W- - - - -N- - - - -KKG- - - - - 278
Db 1928 PEVLPGFRFSGHSPRVDHVSVRGKNHVSMAQVRIIVLEAVSGLOVPNCCEPGIATG 1987
QY 279 - - - - -NP- - - - - - - - - - -ILS- - - - - -AG- - - - - - 285
Db 1988 TERNFTARVQGRVAVAYFSLQKVQGDLSVLISGRDVTYTFVAAGLLEIQVAFNALG 2047
QY 286 - - - - -V-D- - - - -K- - - - -T- - - - -T- - - - -I- - - - -WD- - - - - 294
Db 2048 SENRTLVEQDAVQVVALQSGCFNRSQAFAATSPSPRRVAYHWDGDSGPGQDTBE 2107
QY 295 - - - - -A-HT- - - - -G- - - - -E- - - - -A- - - - -KQ- - - - -Q- - - - -FP- - - - - 305
Db 2108 PRAEHSVLRGDRVQVNASNLVSFFVAQATVTVQVLACREPEVDVVLPLQVLMRSQRN 2167
QY 306 F- - - - -H- - - - - - - - - - -SA- - - - -P- - - - -AL- - - - -DV- - - - - 314
Db 2168 YLEAHVLDLDCVTYQTEYRWEVYRTASQCPGRPARVALPGVDVSRPLVLRALPVGH 2227
QY 315 - - - - -D- - - - - - - - - - - - - - - -W- - - - - - - - - - -QS- - - - -N- - - - - 319
Db 2228 YCFVVFVSGDTPLTOSIQANVTVAERLVPPIIEGGSYRVMSDTRDLVLDGSESYDNLE 2287
QY 320 - - - - -N- - - - - - - - - - - - - - - -TF- - - - - - - - - - - 322
Db 2288 DGDQTPLSFHWACVASTQREAGCALNFGPRGSSVTIPIERLAAGVEYTFSLTVWKAGR 2347
QY 323 - - - - -A- - - - - - - - - - -SC- - - - - - - - - - - - - - - - 325
Db 2348 KEEATNQTVLIRSGRPVIVLEVCCKAQAVVEVRSVYVLEGRCLNCSGSKRGWAA 2407
QY 326 - - - - -S-T- - - - -D- - - - - - - - - - -M- - - - - - - - - - -C- - - - -I- - - - - 331
Db 2408 RTFSNKTLVLDETTTSGAGMRLVLRGRVLRDGBGYTFTLVLRSGSEEGCASIRLSP 2467
QY 332 - - - - -H- - - - - - - - - - -V- - - - - - - - - - - - - - - -C- - - - - 334
Db 2468 NRPLGSCRLPLGVAHALTTKVHPECTGCHDAEDAGAPLVYALLRRRCROCHCEFCV 2527
QY 335 - - - - -K- - - - - - - - - - -LG- - - - - - - - - - - - - - - - 337
Db 2528 YKGLSSYGAVLPFGPRPHFVGLAVVQDLGAADVVALNRSLAITLPEPNGSATGLTVW 2587
QY 338 - - - - -Q-D- - - - - - - - - - - - - - - -R- - - - - - - - - - -P- - - - - 341
Db 2588 LHGLTASVLPGLLRQADPOHVIEYSALVTVLNEYERALDVAABPKHERQRAQIRKNIT 2647
QY 342 - - - - -I- - - - - - - - - - -K-T- - - - - - - - - - -FO- - - - - - 346
Db 2648 ETLVSLRVTVDDIQIIAALAQCMGSPRELVCRLCKQTLHKLEAMMLILQAEATTAGTV 2707
QY 347 - - - - -G- - - - -H- - - - - - - - - - -T- - - - - - - - - - - 349
Db 2708 TPTAIGDSLINITGDLIHLASSDVRAPQPSSELGAESPMSRWASQAYNLTSALMILMRSR 2767
QY 350 - - - - -NE- 351
Db 2768 VLNEEPLTLAGEIIVAGKRSRPSRLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIF 2827
QY 352 - - - - -V- - - - -N- - - - - - - - - - -AI- - - - -K- - - - -W- - - - - 357
Db 2828 LVDSNPPFGYISNYTVSTVKVASMAFQOAGAQIPIERLASERAITVKVPNNSDWAARGH 2887
QY 358 - - - - -D- - - - -P- - - - -TG- - - - -N- - - - - - - - - - -LLA- - - - - 365

Db 2888 RSSANSANVVQPOQASVGAIVTLDSSNPAAGLHLQNLNTYLLDGHYLSBEEPEPYLAVYLH 2947
QY 366 S- - - - -CS- - - - - - - - - - -D- - - - - - - - - - -D- - - - - 370
Db 2948 SEPRPNEHNCASRRIRPESLOGADHRPYTFFISPGSRDPAGSYHLNLSHFRMSALQVS 3007
QY 371 - - - - -M-T- - - - -L- - - - -KIM- - - - -S- - - - -MK- - - - -Q- - - - - 380
Db 3008 VGLYTSLCQYFSEEDVMWRTGELLPLEETSQPAVCLTRHLTAFGASLFPVPPSHVRFPV 3067
QY 381 - - - - -D-N- - - - - - - - - - -C- - - - -V- - - - -H- - - - -DLQ- - - - -QNN- - - - -K- - - - -EYTIK 398
Db 3068 EPTADVNIYMLTCAVCLVTYVMMAILHKLJD-QLDASGRAIPFCGQGRFYEIL-VK 3125
QY 399 - - - - -W- - - - -S- - - - -PT- - - - - - - - - - -G- - - - - - - - - - -P- - - - -GT- - - - - 406
Db 3126 TCGRGSGTTAHVGMLYGVDSRSGHRHLDGDRAFHRNSLDIFRIATPHSLGSSWVKIRVW 3185
QY 407 - - - - -NN- - - - -P- - - - - - - - - - -N- - - - - - - - - - -AN- - - - -L- - - - -M- - - - -L- - - - -AS- - - - - 419
Db 3186 HDNKGLSPAFLQHVIVRDLQTARSAPFLVNDWLSVETEANGGLVEKEVLAASDAALLRF 3245
QY 420 - - - - -FD- - - - - - - - - - -S- - - - -T- - - - -V- - - - -R- - - - -L- - - - -W- - - - -D- - - - - 428
Db 3246 RRLVAELQRFQKHIMLSIWDPRPSRFTRIQATCCVLLICLFLGANAVMYGAVGDS 3305
QY 429 - - - - -VD- - - - - - - - - - - - - - - -R- - - - - - - - - - -G- - - - - 432
Db 3306 AVSTGHVSRSLPSLVDTVAVGLVSSVVVYVYLAILFLPRMSRKVAGSPSPAGQOVL 3365
QY 433 - - - - -I- - - - -C- - - - - - - - - - -H- - - - - - - - - - -TLT- - - - - 439
Db 3366 DIDSCLDSSVLDSFLTFTSGLHAEQAFVQOMKSDLPDDSKSLVCWPSGEGTILSWPDLLS 3425
QY 440 - - - - - - - - - - - - - - - -K- - - - - - - - - - - - - - - - 440
Db 3426 DPSIVGSNLRQLARGOAGHGLGPEEDGFLSPSPAKSPASFSASDEDLIOQVLAEGVSSPA 3485
QY 441 - - - - -H- -Q- - - - - - - - - - -E- - - - - 443
Db 3486 PTQDTHMETDLSLSSSTPGEXTETLALQRLGELPPSPGLNWEQQAARLSRTGLVEGL 3545
QY 444 - - - - -P- - - - - - - - - - -V- - - - -Y- - - - - - - - - - -SVA- - - - - - - - - - - 449
Db 3546 RKRLLPANCASLAHGLSLLLVAVAVSGWVGASFPFGVSVAVMLSSSSASFLASFLGWEP 3605
QY 450 - - - - -FS- - - - - - - - - - -PD- 453
Db 3606 LKVLLEALYFSLVAKRLHPDEDDTLVESPAVTPVSARVPRVPPHGFALFLAKEEARKVK 3665
QY 454 - - - - -G- - - - -R- - - - -Y- - - - -LAS- - - - -GSFD- - - - -KC- - - - - - - - - - -H- - - - -I- - - - - 468
Db 3666 RUHGMRLSLLVYMLFLLVTLTASYG--DASCHGHAYRLQSAIKQELHSAFLAITRSEEL 3723
QY 469 - - - - -W- - - - - - - - - - -N- - - - -T- - - - -OV- - - - - - - - - - -C- - - - - - - - - - -L- - - - - 475
Db 3724 WFWMAHVLLPYVHGQSSPELQFPRLRQVRLQEAALYDPDPPGPRVHTCSAAGSFSTSDYDV 3783
QY 476 - - - - -H- - - - - - - - - - -Y- - - - -L- - - - - - - - - - -NG- - - - -Q- - - - - - - - - - -VL- - - - -L 484
Db 3784 GWESPHNGSGTWAYSAPDLLLGAWSWGSCAVYDGGVQELGSLSEBSRDLRPLQLHNWL 3843
QY 485 - - - - -N- - - - - - - - - - -L- - - - -GR- - - - - - - - - - - - - - - - - 488
Db 3844 DNRSAVLELTRYSPAVGLHAAVTLRLEFFPAAGALAAALSVRPFALRRLSAGLSLPLLT 3903
QY 489 SICL- - - - - - - - - - -Y- -L- - - - - 495
Db 3904 SVCLLLFAVHFAVEARTWHREGVRVRLGAWARLLVALTAATALVRLAQIQAADROW 3963
QY 496 - - - - -P- - - - - - - - - - -H- -H- - - - - 498
Db 3964 TRFVRGRPRRFTSFQVAHVSSAARGLAASLLFLLLVKAAQHVRFVRQWSVFGKLCRAL 4023

QY 499 -----LVV-----I----- 502
Db 4024 PELGVTGLVGLVAVQAAILLVSSCVDSLSVQAALLVCPGTGLSTLCPAESWHL 4083
QY 503 PL--V-----AL--I-----ELL-----Y-----L 513
Db 4084 PLLCVGLWALRLWGLRGLGAVILWRVYHALRGELYRPAWEPQDYEMVELFLRLRLWML 4143
QY 514 -K 514
Db 4144 SK 4145

RESULT 15

US-09-052-469-8
; Sequence 8, Application US/09052469
; Patent No. 6380360
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052.469
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-052-469-8

Query Match 69.2%; Score 2526.5; DB 3; Length 4302;

Best Local Similarity 9.9%; Pred.No.1.5e-31;
Matches 405; Conservative 13; Mismatches 13; Indels 3575; Gaps 351;

QY 1 MSIS-----S-----D-----EVN-----F----- 10

Db 72 LDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEGIPANLFNLSEINLSGNPPECDC 131
QY 11 -LVY-----R-----Y----- 15
Db 132 GLAWLPRWAEQQVRVVQPEAATCAGPGSLAGPQLLGIPLLDSCGEEYVACLPDSSGT 191
QY 16 -----LQ-----E-----SG- 20
Db 192 VAAVSFAAHEGLLOPEACSAFCSTGQGLAALSEQWCLCGAAQSSASFACLSLCSGP 251
QY 21 -----F-S-----H-S-----AF-----T-FG----- 29
Db 252 PPPAPTCTGPTLLQHVFPASPGATLVGPHGLASQGLAAFAHIAAPLPVTAATWDFDGS 311
QY 30 --I-----KSH-----I-----S-QSN- 38
Db 312 AEVDAAGPAASHRYVLPGRYHVTAVALGAGSALLGTQVVEAAPAAALELVCPSSVQSD 371
QY 39 -----I-N-----G-----AL-----V-----P----- 46
Db 372 SLDLSTQNRGGGLEAAYSIVALGEEPARAVHPPLCPSTDTEIFPGNGHCYRLVVEKAAMLQ 431
QY 47 -----AA-----LIS-----I-----IQ--K-----G-----L-- 57
Db 432 AQEQCQAWAGALAMVDSPAVQRFVLSRVTRSLDVWIGFSTVQGVVEGPAQGEAFSLES 491
QY 58 -Q-----YV-E-----AE----- 63
Db 492 CQNLPGEPHPATAEHCVRGLGPTGMCNTDLCAPHSYVCELOPPGPGVQDAENLLVGAPSG 551
QY 64 -----V-----S-----I-- 66
Db 552 DLQGLTPLAQDGLSAPHEPVEVMVFPGLRLSREAFLTABFGTOELRRPQLRLQVTR 611
QY 67 -----N-----E-----D-----G-T- 71
Db 612 LLSTAGTPENGSEPESSPDNRTQLAPACMPGGRWCPCGANICLPDASCHPQACANGCTIS 671
QY 72 -----LF-----D----- 74
Db 672 GPGLPGAYALWREFLFSVAAGPPAQSVTLHGQDVLMLPGDLVGLQHDAGPGALLHCSP 731
QY 75 -----G-RP----- 77
Db 732 APGHPOGPAPYLSANASSWLPPLPAOLEGTWACPCACALRLLAATEOLTVLLGLRPNPGLR 791
QY 78 -----I-E-----LS-----L-I-----D- 85
Db 792 MGRGYEVRAEVGNVGRHNLSCSFDVVSFVAGLRVIYPAPRDGRLVVPNTGSAVLQVDS 851
QY 86 -----A-----V-MP-----DV 91
Db 852 GANATATARWPGSVSARFENPCALVATFVPGCPMETNDTLFVVVALPWLSEGEHVVDV 911
QY 92 V-----O-T-----R-----Q-Q-----A-Y-----R-----D 101
Db 912 VVENSASRANLSLRVTAEPIGCLRATPSPEARVLQGLVRYSPVVEAGSDMVFRWTIND 971
QY 102 KLAQ-----Q-----QAAAA-----AAA-----AS-----Q-Q----- 121
Db 972 K--QSLTFQNVVFNVIYQSAAVFKLSLTASNHVSNTVNVYTVRMMNRMQGLQVSTVPA 1029
QY 122 -----G-----SA-K-----N-----GE-----N-----TA----- 131
Db 1030 VLSPNATLVLTGGVLVDSAVEAFLWNFGDGEQALHQFQPPYNESFPVDPDSVAQVLVBH 1089
QY 132 N-----GE-----EN-----G--A-----H 139
Db 1090 NVMHTYAAEGEYLLTTLASNAPENLTQQVPVSVRASLPVAVGVSDGLVAGRPVTFYPH 1149
QY 140 -----T-----I-----ANNHT----- 146

Db 1150 PLSPGGVLTWDFDGSFVLTSQSPAANTHYASRGTYHVRLEVNNTVSGAAQAQVRVF 1209
QY 147 ---DM-M-E---V---D-GD--- 154
Db 1210 EELRGLSDMSLAVEQCAPVVSAVOTGDNITWTFDMGDTVLSPGEATVEHVYLAQN 1269
QY 155 ---VE---IP-P---NKA--- 162
Db 1270 CTVTGAASPAHLARSLHVLVFLVLEVRPEAPACIPTQDARLTAVVTGNPAHYLPDWT 1329
QY 163 ---V---L-R-G---H---ESV--- 172
Db 1330 FGDSSNTVRCPTVTHNFRSGTFPLALVLSRVNRAHYFTSICVEPEVGNVTLOPER 1389
QY 173 -FI---CAW---NP-VS---D-LL---AS- 186
Db 1390 QFVLGDDEAWLVACANPPFPYRYTWDGTEEAAPTRARGEVTFIYRDPGSYLVTASN 1449
QY 187 ---GS---GD--- 190
Db 1450 NISAANDSALVEQEPVLVTSIKVNGSLGLELOQPYLFSAVGRPAASYLDLGDGWL 1509
QY 191 ---S---TARI-WN-L-SE---N---S 202
Db 1510 GPEVTHAYNSTGDTVRVAGMNEVSEAWLNVTVRRVRGLVNVASRTVVPLNGSVSFS 1569
QY 203 TS---GS-TOL-VL-RHC---I---RE-G---G-OD--- 221
Db 1570 TSLEAGSDVRSVWLCOR-CTPIPGGPTISYTRSVGTENIIVTAENEVGSQDSIFVYV 1628
QY 222 ---V---PSN---KD---V---T---SL--- 231
Db 1629 LQIEGLQVVGGRYPFTNHTVQLQAVVRDGTNVSYSWTAWRDRGPALAGSGKFSITVL 1688
QY 232 ---D---W---N--- 234
Db 1689 EAGTYHVQLRATNMLGSAWADCTMDFEVPVGLMTVATSPNPAAVNTSVTSLAELAGSGV 1748
QY 235 ---S-EG---T-L-LAT---GS---YD--- 246
Db 1749 VYTWSLEGLSWTSEBPTTHSPFTGLHLVTWTAGNPLGSANATVEVDVQVPVSLSIR 1808
QY 247 ---G-F-A---R---I---W---TK---D-G---NL 258
Db 1809 ASEPGGSFVAAGSVFPWGLATGTVNWCMAVPGGSKRGPHVTWVPDAGTFSIRLN- 1867
QY 259 AS---T---L---QO-H---K---G--- 267
Db 1868 ASNAVSWSATYNLTAEPIVGLVLWASSKVAPGQLVHFQIILAAAGSAVTFRLQVGGAN 1927
QY 268 P-I---F-AL---K---W---N-KKG--- 278
Db 1928 FEVLGPRFHSPPRVGDHVVSVRGKNHVSQAQVRIVVLEAVSGLQMPNCCPEGIATG 1987
QY 279 ---NF---ILS---AG--- 285
Db 1988 TERNFTARVQGRSAVAYFSLQKQDLSVLISGRDVTYTPVAAGLLEIQVRAFNALG 2047
QY 286 ---V-D---K-T---T---II-WD--- 294
Db 2048 SENRTLVLVEQDAVQVVALQSGPCFTNRSQAPEAATSPSPRVAYHWDGDSGPGQDTDE 2107
QY 295 ---A-HT---G-E-A---KO---Q---FP--- 305
Db 2108 PRAEHSYLRGDRYRVQVNASNLVSFFVAQATVTVQLACREPEVDVVLPLQVLMRSQRN 2167
QY 306 F---H---SA---P---AL---DV--- 314
Db 2168 YLEAHVDLRDCTVYQTYEYRWEVYRTASCQBPGRPARVALPGVDVSRPRLVPLALPVGH 2227
QY 315 ---D---W---QS---N--- 319
Db 2228 YCFVFWVSGDTELTSIQANVTVAPELVRPLPIEGGSYRVWSDTRDLVLDGSESYDPNLE 2287

QY 320 ---N---TF--- 322
Db 2288 DGDQTPLSFWHACVASTQREAGCALNFGPRGSSTVTIIPRERLAAGVEYTFSLTVMWKA 2347
QY 323 ---A---SC--- 325
Db 2348 KEBATQTVLIRSGRPVIVLESCVCKAQAVYEVSRSSVYVLEGRCLNCSGSKRGRWAA 2407
QY 326 ---S-T---D---M---C-I--- 331
Db 2408 RTEFNKTLVLDETTTSGAGMRLVLRGVLRDGEGYTFTLTVLGRSGEBEGCASIRLSP 2467
QY 332 ---H---V---C- 334
Db 2468 NRPLGSGCELPLGAVHALTKVHECTGCHDAEDAGAPLVYALLRLRCRQCHCEFCV 2527
QY 335 -K---LG--- 337
Db 2528 YKGLSSYGVAVLPFGPRPHEVGLAVVQDQCGAAVVALNRSIAITLPEPNSATGLTVM 2587
QY 338 ---Q-D---R---P--- 341
Db 2588 LHGLTASVLPGLLRQADPOHVIEYSALVTVLNEYERALDVAAEPKHERQHRQAIRKNI 2647
QY 342 ---I---K-T---FO--- 346
Db 2648 ETIVSLRVHTVDDIQOIAAALACQMPSELVCRSCLKQTLHLKLEAMMLILOAETTAGTV 2707
QY 347 ---G-H---T--- 349
Db 2708 TPTAIGDSILNITGDLHLASSDVRAPOPSSELGABSPSRMVASQAVNLSALMRLMRSR 2767
QY 350 -NE--- 351
Db 2768 VLNEEPLTLAGEEIVAQKRSDPRSLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIF 2827
QY 352 -V---N---AI-K---W--- 357
Db 2828 LVDSNPPFGYISNYTVSTKVASMAFOTQAGAQIPIERLASERAITVKVPNNSDMAARGH 2887
QY 358 ---D-P-TG---N---LLA--- 365
Db 2888 RSSANSANVVQPQASVGAVVTLDDSSNPAAGLHLQNLVTLDDGHVLSBEPEPYLAVLH 2947
QY 366 S---CS---D--- 370
Db 2948 SEPRNEHNCASRRIRPESLQCADHRPYTFFISPGSRDPAGSYHLNLSHFRWSALQVS 3007
QY 371 -M-T-L---KIW-S---MK---Q--- 380
Db 3008 VGLYTSLCQYFSEEDMWRTGELLPLEETSPPQAVCLTRHLTAFGASLFPVPPSHVRFPVP 3067
QY 381 ---D-N---C---V---H-DLQ---OHN-K-EIYTIK 398
Db 3068 EPTADVNIYVMTCAVCLVTVYVMAAILHKLD-QLDASGRAIPFCQGRGRFKYEIL-VK 3125
QY 399 -W---S-PT---G---P---GT--- 406
Db 3126 TGMGRSGTTAHVGMILYGVDSRSGHRLDGDRAFHNSLDIFRIATPHSLGSMVKIRVM 3185
QY 407 -NN---P---N---AN-L-M-L-AS-AS- 419
Db 3186 HDNKGSLPAWFIQHVIVRDLQTAARSFAFFLVNDWLVSVEANGELVKEVLAASDAALLRF 3245
QY 420 ---FD---S-T-V-R---L---W---D- 428
Db 3246 RRLVVAELQGRFQKHILWLSIWDPRPSRFTRIQATCCVLLJLCLFLGANAVYVAGVDS 3305
QY 429 ---VD---R---G--- 432
Db 3306 AYSTGHVSRLSPLSDVTAVGLVSSVVVYVYLAILFLFRMSRSKVAGSPSPAGQOVL 3365

QY 433 -I--C---I---H-----TLT----- 439
Db | | :
3366 DIDSCLDSSVLDSSFLTFSLHAEAFVGQMKSDLFLDDSKSLVCWPSGEGTILSWPDLSD 3425
QY 440 -----K----- 440
Db | |
3426 PSIVGSLNQLARGQAGHGLGPEEDGFSLASPYSPAKSFASDEDLIQOVLAEVGSPPAP 3485
QY 441 ---H-----O-----E--- 443
Db | |
3486 TQDTHMETDLLSSLSTPGKTEKTLALQRLGELGPPSPGLNWEQQAARLSRTGLVEGLR 3545
QY 444 ---P-----V--Y-----SVA----- 449
Db | | :
3546 KRLLPAWCAKSLANGLSLLVAVAVAGVGMVGASFPQGVSAWLLSSSASFLASFLGWEPL 3605
QY 450 ---FS-----PD----- 453
Db | |
3606 KVLLEALYFSLVAKRLHPDDEDTLVESPAVTPVSARVPRVPPHGFALFLAKEARKVKR 3665
QY 454 --Q--R--Y-----LAS--GSPD--KC-----V--H-----I----- 468
Db | | :
3666 LHGMLRSLLVYMLFLVLVTLASYG--DASCHGAYRLQSAIKOELHSRAPLAITRSEELW 3723
QY 469 -W-----N--T-----QV-----C-----L--- 475
Db | | :
3724 PWWAHVLLPYVHGNQSSPELGPRLQVRLQALYDPDPGPRVHTCSAAGGFSTSDYDVG 3783
QY 476 ---H-----Y-----L-----NG--Q-----VL-----L- 484
Db | | :
3784 WESPHNGSGTWAYSAPDLLGAWSGWCAVYDSGGYVQELGSLSESRDLRFLQLHNWLD 3843
QY 485 N-----L-----GR-----S 489
Db | |
3844 NRSRAVLELTRYSPAVGLHAAVTLRLEPPAAGRALAALSVRPFALRLRLSAGLSPLITS 3903
QY 490 ICL-----Y-----T-----L----- 495
Db | | :
3904 VCLLLFAVHPAVAEARTWHREGWRVRLGAWARWLLVALTAATLALVRLAQLGAADQWT 3963
QY 496 ---P-----H-----H----- 498
Db | | :
3964 RFVGRPRRFTSPDQVAHVSSAARGLAASLLFLLLVKAAQHVRVFRQWSVFGKTLCLALP 4023
QY 499 ---LW-----I-----P 503
Db | | :
4024 ELGVTGLVGLVAVAYQAAILLVSCVDSLWSVAQALLVLCPTGLSTLCPAESWHLSP 4083
QY 504 L--V-----AL-----I-----ELL-----V-----L- 513
Db | | :
4084 LLCVGLWALRLWGLRLGAVILRWRYHALRGELYRPAWEPQDYEMVELFLRLRLWGLS 4143
QY 514 K 514
Db |
4144 K 4144

Search completed: January 3, 2005, 15:29:32
Job time : 101.667 secs

Db 172 ATAAATTSAGVSKQNSKREAVTNGENRAHSV-NNHA-KPMEIDGVEIPSSKATVLR 229
QY 167 GHESEVFIICAWNVPVSDLLASGSDSTARIWNLSNSTSGSTQVLVLRHCIREGQDVPSNK 226
Db 230 GHESEVFIICAWNVPVSDLLASGSDSTARIWNLSNSTSGSTQVLVLRHCIREGQDVPSNK 289
QY 227 DVTSLDWNSGTTLATGSDYDGFARIWTKDGNLASTLGOHKGPALKNKKNFIILSAGV 286
Db 290 DVTSLDWNTGTTLATGSDYDGFARIWTKDGNLASTLGOHKGPALKNKKNFIILSAGV 349
QY 287 DKTIIWDHTAGTGAQKQFPFHSAPALDVWQSN--TFASCSSTDMCHVCKLGDRIKTF 345
Db 350 DKTIIWDHTAGTGAQKQFPFHSAPALDVWQ--NNTTFASCSSTDMCHVCKLGDRIKTF 408
QY 346 QGHTNEVNAIKWPTGNLLASCSDMTLKIWSMKQNCVHDLQHNKEIYTIKWSPTGPG 405
Db 409 QGHTNEVNAIKWPTGNLLASCSDMTLKIWSMKQNCVHDLQHNKEIYTIKWSPTGPA 468
QY 406 TNNPNANMLASFSFVSLVLDVDRGICHTLTKHOEPVVSFAFSDGKYLASGSDK 465
Db 469 TSNPNSNIMLASFSFVSLVLDVDRGICHTLTKHOEPVVSFAFSDGKYLASGSDK 528
QY 466 VHIWNTQ---VCLH-Y--LNG--QVLLN-----LGR-----V-L-D---L-RK 506
Db 529 VHIWNTQSNLV--HSYRGTTGIFVWCNARGDKVGASDGSVC-----V- 572
QY 507 LIELLV-K 514
Db 573 L-D---L-RK 577

RESULT 2

US-10-264-049-2892
; Sequence 2892, Application US/10264049
; Publication No. US2004005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2892
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2892

Query Match 84.0%; Score 3064.6; DB 15; Length 542;
Best Local Similarity 73.2%; Pred. No. 1.9e-43;
Matches 416; Conservative 45; Mismatches 23; Indels 84; Gaps 27;
QY 5 SDEVNF---LV-----YR-----YLOESGFSHAFPGIKSHISQSNING 41
Db 1 TDC--FKESKVDLQSQKPLRKTHTLCGCTGFLTC-GFSHSAFTPIESHSISQSNING 57
QY 42 ALVPPAALISITQKGLQVAAEVASINEDGTLFDGRPIESLSLIDAVNPVQVTRQAYRD 101
Db 58 TLVPPAALISITQKGLQVAAEVASINEDGTLFDGRPIESLSLIDAVNPVQVTRQAFRE 117
QY 102 KLAQQQ--AAAAA--AAASQ-----QG-S-----AKNGENTANGENGAAHTTANNHTDM 148
Db 118 KLAQQQASAAAAAATAAATAATTSAGVSHQNPKNREATVNGEENRAHSV-NNHA-K 175

QY 149 -MEVDGVEIPPNKAVLVRGHESEVFIICAWNVPVSDLLASGSDSTARIWNLSNSTSGST 207
Db 176 PMEIDGVEIPSSKATVLRGHESEVFIICAWNVPVSDLLASGSDSTARIWNLSNSTSGST 235
QY 208 QVLVLRHCIREGQDVPSNKDVTSLDWNSGTTLATGSDYDGFARIWTKDGNLASTLGOHKG 267
Db 236 QVLVLRHCIREGQDVPSNKDVTSLDWNTGTTLATGSDYDGFARIWTKDGNLASTLGOHKG 295
QY 268 PIFALKKNKKNFIILSAGVDTKTTIIWDHTAGTGAQKQFPFHSAPALDVWQSN--TFASCS 326
Db 296 PIFALKKNKKNFIILSAGVDTKTTIIWDHTAGTGAQKQFPFHSAPALDVWQ--NNTTFASCS 354
QY 327 TDMCHVCKLGDRIKTFQGHTEVNAIKWPTGNLLASCSDMTLKIWSMKQNCVHDL 386
Db 355 TDMCHVCKLGDRIKTFQGHTEVNAIKWPTGNLLASCSDMTLKIWSMKQNCVHDL 414
QY 387 LQHNKEIYTIKWSPTGPTGNPNANMLASFSFVSLVLDVDRGICHTLTKHOEPVY 446
Db 415 LQHNKEIYTIKWSPTGPTGNPNANMLASFSFVSLVLDVDRGICHTLTKHOEPVY 474
QY 447 SVAFSPDGKYLASGSDKCVHIWNTQ---VCLH-Y--LNG--QVLLN-----LGR--- 488
Db 475 SVAFSPDGKYLASGSDKCVHIWNTQSNLV--HSYRGTTGIFVWCNARGDKVGASD 532
QY 489 -SICLYTLPHLWVPLVALIELLV-K 514
Db 533 GSVC-----V-L-D---L-RK 542

RESULT 3

US-10-363-829-446
; Sequence 446, Application US/10363829
; Publication No. US20040142331A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;
; APPLICANT: Altus, Christina M.; Dufour, Gerard E.;
; APPLICANT: Chalup, Michael S.; Jackson, Jennifer L.;
; APPLICANT: Jones, Anisea L.; Yu, Jimmy Y.;
; APPLICANT: Wright, Rachel J.; Gietzen, Darryl;
; APPLICANT: Liu, Tommy F.; Yap, Pierre E.;
; APPLICANT: Dahl, Christopher R.; Momiyama, Monika G.;
; APPLICANT: Bradley, Diana L.; Rohatgi, Sameer D.;
; APPLICANT: Harris, Bernard; Roseberry Lincoln, Ann M.;
; APPLICANT: Gerstein, Jr., Edward H.; Peralta, Careyna H.;
; APPLICANT: David, Marie H.; Panzer, Scott R.;
; APPLICANT: Flores, Vincent Z.; Daffo, Abel;
; APPLICANT: Marwaha, Rakesh; Chen, Alice J.;
; APPLICANT: Chang, Simon C.; Au, Alan P.;
; APPLICANT: Inman, Rebekah R.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PT-1183 USN
; CURRENT APPLICATION NUMBER: US/10/363,829
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: PCT/US01/27628
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/229,751
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,749
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,750
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,747
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,748
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,583
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,517
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,610
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,597

; PRIOR FILING DATE: 2000-09-06
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: PERL Program
; SEQ ID NO 446
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: LI:021759.1.orf2:2000SEP08
US-10-363-829-446

Query Match 75.2%; Score 2744; DB 16; Length 395;
Best Local Similarity 99.5%; Pred. No. 4.3e-38;
Matches 369; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSISSDEVNPLVRYLOESGFSAFTFGIKSHISQSNINGALVPPAAALISIIQKGLQYV 60
DB 5 MSISSDEVNPLVRYLOESGFSAFTFGIESHISQSNINGALVPPAAALISIIQKGLQYV 64
QY 61 EAEVSI NEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQOQAAAAAASQ 120
DB 65 EAEVSI NEDGTLFDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQOQAAAAAASQ 124
QY 121 QGSAGKNGTANGENGAGHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVEFICAWN 180
DB 125 QGSAGKNGTANGENGAGHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVEFICAWN 184
QY 181 SLLASGSGDSTARIWNLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
DB 185 SLLASGSGDSTARIWNLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244
QY 241 ATGSYDGFARIWTKDGNLSTLGOHKGPFPALKWKKGNFILLSAGYDKTTIWDATGEA 300
DB 245 ATGSYDGFARIWTKDGNLSTLGOHKGPFPALKWKKGNFILLSAGYDKTTIWDATGEA 304
QY 301 KQFPFHPSPALPDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQCHTNEVNAIKWDPT 360
DB 305 KQFPFHPSPALPDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQCHTNEVNAIKWDPT 364
QY 361 GNLASCSDDMTLKIWSMKQDNCVHDLQAHN 391
DB 365 GNLASCSDDMTLKIWSMKQDNCVHDLQAHN 395

RESULT 4

US-10-451-168-78
; Sequence 78, Application US/10451168
; Publication No. US20040091969A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50039
; CURRENT APPLICATION NUMBER: US/10/451,168
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: PCT/US01/49232
; PRIOR FILING DATE: 2000-12-17
; PRIOR APPLICATION NUMBER: 60/256,710
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/257,048
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/260,482
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/264,922
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/266,797
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/276,988
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/281,535

; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/289,622
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 5635
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-451-168-78

Query Match 71.3%; Score 2601.6; DB 15; Length 5635;
Best Local Similarity 9.8%; Pred. No. 2.1e-29;
Matches 431; Conservative 67; Mismatches 13; Indels 3904; Gaps 372;

Db 841 VSSISQRTGALFILNLWASDKGYICEAENOFKIQOSETTVTVTGLVAPLIGISPVAN 900
 QY 96 --QO-- --A-- --Y-- --R-- --D-- --KL-- 103
 Db 901 VIEGQOULTPCTLLAGNPPIERWIKNSAMLLONPYITVRSOGLHIERVOLQDGGEYTC 960
 QY 104 --A-- --Q-- --Q-- --AA-- 109
 Db 961 VASNAVGTNNKTSVVVHVHFTIHHQOILSTIEGIPVTLPCAKSGNPKPSVIMSKKGL 1020
 QY 110 --AAA-- --A-- --A-- --A-- 118
 Db 1021 ISTSSAKFSAGADSLYVSPGGESEYVCTATNTAGYAKRKVQTVVVRPFGDQRG 1080
 QY 119 --SO-- --A-- --Q-- --GS-- 123
 Db 1081 LSQOKPEISVLAGEEVLTPCEVKSLLPPPIITWAKETQLISPPSPRHTFLPSGMMKITET 1140
 QY 124 --A-- --A-- --K-- --N-- 126
 Db 1141 RTSDSGMYLCVATNIAGNVTAQVAVKLVNHHVPPKIQRGPKHLKVQGVQORVDIPCNAGOTPLP 1200
 QY 127 --GE-- --N-- --T-- --A-- --N-- --G-- 133
 Db 1201 VITWSKGGSTMLVGEHHVSNPDGTLISIDQATPSDAGIYTCVATNIAGTDETEILHVQE 1260
 QY 134 --E-- --N-- --GA-- --H-- 139
 Db 1261 PPTVEDLEPPYNTTFQERVANQRIEPPCPAKGTPKTIKWLHNGRELTGREPGISILEDG 1320
 QY 140 T--IA-- --N-- --N-- --HT-- --DM-- 148
 Db 1321 TLLVIASVTPYDNGEYTCVAVNEAGTTERKYNLKVHVPVVKDKBQTVNVSVLLNQLTNL 1380
 QY 149 --M-- --E-- 150
 Db 1381 FCEVEGTPSIIIMYKDNVQVTSSTIQTWNGKILKLFRATPDAGRYSCAKAINIAGTS 1440
 QY 151 --VD-- --G-- --DV-- --E-- --I-- --P-- --PN 160
 Db 1441 QKYFNIDVLVPPPTIIGTFNPNVSVLNRDVALEQVKGTPFPDIHFKDGKFLGDPN 1500
 QY 161 --K-- --K-- --V-- --V-- 165
 Db 1501 VELLDRGOVLHKNARNNDKRYOCTVSNAGKQAKDKIULTIYNPPSIKGNVTDISVL 1560
 QY 166 --RG-- --H-- 168
 Db 1561 INSILKLECETRGLPMPALITWYKDGQPIWSSQALYIDKGQYLHI PRAQVSDSATYTCVH 1620
 QY 169 --E-- --S-- --VPI-- --C-- --A-- --W-- 177
 Db 1621 ANVAGTAESPHVDVYVPPMIEGNLATPLNKQVVIASHLTLECKAAGNPPSILTLWKDGV 1680
 QY 178 --N-- --P-- --V-- --SD-- 182
 Db 1681 PVKANDNIRIBAGGKLEIMSAQEI DRGOYTCVATSVAGEKEIKYEVDVLVPPAIEGDE 1740
 QY 183 --LL-- --ASGS-- --G-- --D-- --S-- 191
 Db 1741 TSYFIVMNNLELDCHVTGSPPTIMWLKDGQILIDERDGFKILLNGRKLVIQAQVSN 1800
 QY 192 --TA-- --R-- --I-- --W 196
 Db 1801 GLYRCMAANTAGDHKKFEVTVHVPPTIKSSGLSERVVVKYPVALQCIANGIENPSITW 1860
 QY 197 --NL-- --S-- 199
 Db 1861 LKDDQPVNTAQGNLKIQSGRVLQIAKTLLLEDAGRYTCVATNAAGETQHQIQLHVHPPS 1920
 QY 200 --E-- --N-- --S-- --SGST-- 207
 Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNFPVPTIWKDNRLLSGSTSMTFLNRGQIIDIES 1980

QY 208 --Q-- --L-- --V-- --LR-- --H-- --C-- --I-- 215
 Db 1981 AOISDAGIYKCVAINSAGATELFYSLQVHVHAPSISGSNNMVAVVVNVPRLCEARGIPA 2040
 QY 216 --R-- --E-- --G-- --Q-- --D-- 221
 Db 2041 PSLTWLKDGSVPSSFNGLQVLSGRILALTSAQISDTRGYTCVAVNAAGEKQORDIDLVR 2100
 QY 222 --VPSN-- --KD-- --VT-- --S-- --L-- 231
 Db 2101 YVPPNIMGEEQNVLSISQAVELLQCSDAIPPTLTWMLKDGHPLLKKPGLSISNRSVLK 2160
 QY 232 --D-- --W-- --N-- --S-- --EG-- --TLL-- --A 241
 Db 2161 IEDAQVQDTRGYTCVATNAGTEKNVNNVWPNIGGSDDELTLQTLVIEGNLSLICES 2220
 QY 242 T--GS-- --Y-- --D-- --G-- --F-- --A-- 249
 Db 2221 SGIPPPNLIWKKGSPVLTDSMGRVRLISGGRQLQISIAEKSDAALYSCVASNAGTAKK 2280
 QY 250 --R-- --I-- --WKDG-- 256
 Db 2281 EYNLQVYIRPTITNSGSHPTIIVTRGKSI SLECEVOGIPPTVTWMDKGHPLIKAKGVE 2340
 QY 257 --NL-- --AS-- --T-- --L-- --QO-- 265
 Db 2341 ILDEGHILQLKNIHVSDTRGYCVAVNVAGTMDKKYDLSVHAPPSIIGNHRSPEINISVVE 2400
 QY 266 K--G-- --P-- --I-- --FALK-- --W-- 274
 Db 2401 KNSVSLTCEASGIPLPSITWF--KDGWPVLSNSVRILSGRMLRMQTTMEDAQGYTCV 2458
 QY 275 --N-- --K-- --K-- --GN-- 279
 Db 2459 VRNAAAGEERKIFGLSVLPVPHIVGENTLEDVKVKEKQSVTLTCEVTGNPVPBITWHKQO 2518
 QY 280 --F-- --I-- --L-- --S-- --AGV-- 286
 Db 2519 PIQDEDAHIIISGRFLQITNVQVPHTRGYTCCLASSPAGHKSRPSFLNVFVSPPTIAGVGS 2578
 QY 287 D--K-- --T-- --I-- --I-- --W-- 293
 Db 2579 DGNPEDVTILNSPTSLVCEAYSPPATITWPKDGTPLESNRNIRILPGGRTLQILINAOE 2638
 QY 294 D--A-- --H-- --TG-- --BA 300
 Db 2639 DNAGRYSCVATNEAGEMIKHYEVKVIPIPIINKGDLWGFGLSPKEVKIKVNNLTLECEA 2698
 QY 301 --K-- --Q-- --Q-- --F 304
 Db 2699 YAIPTSASLSWYKDGQPLKSDDHVNIAANGHTLIQKEAQISDTRGYTCVASNIAGEDELDF 2758
 QY 305 --P-- --P-- --H-- 307
 Db 2759 DVNIQVPPSFOKLWEIGNMLDTRNGEAKDVIINNPISYCETNAAPPPTLTWYKDGHP 2818
 QY 308 --SA-- --P-- --A-- --LDV-- --D-- 315
 Db 2819 TSSDKVLIILPGGRVLIQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPIIKGANSDL 2878
 QY 316 --WQ-- --SN-- 319
 Db 2879 PEBVTVLVNSKALIECLSSGPAPRNSWQKQOPLLEDDHHKFLSNRILQILNTOITDI 2938
 QY 320 --NT-- --FAS-- --C-- --S-- 326
 Db 2939 GRVYCVAEANTAGSAKKFVNLNVHVPSPVIGPKSENITVVVNNFISLTCEVSGFPPLSLW 2998
 QY 327 --T-- --D-- --MCI-- 331
 Db 2999 LKNEQPIKLTNTLIVPGGRTLQIIRAKVSDGGEYTCIAINQAGESKKKFLSVYVPPSI 3058

QY 332 --H-----V--C-----K-----L--GQ----- 338
Db 3059 KDHDSLSVVNREGTSVLECSNAVPPVITWYKNGRMITESTHVEIADQMLHIK 3118
QY 339 -----D-----R-----P-IK-----TF----- 345
Db 3119 KAEVSDTGQVCRAINVAGRDDKRNFLNVVPPSIEGPEREVIVETISNPVTLTCDATGI 3178
QY 346 -----Q-----G-----H-----T----- 349
Db 3179 PPPTIAWLKXHKIENSLSLEVRILSGSKLQIARSQHSNGNYTCIASNMGEKAQKYF 3238
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360
Db 3239 LSIQPPSVAGABIPSDSVLLGENVELVCNANGIPTPLIOWLKDGKPIASGETERIRVS 3298
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL----- 373
Db 3299 ANGSTLNIYALTSDTGKYTCVATNPAGEEDRIFNLNVVPTPIRGNKDEAEKMLTVDPT 3358
QY 374 --K-----I-W-----SM-----KO--D--D-----NCV----- 384
Db 3359 SINIECRATGTPPPQINLWLNKGLPLPLSSHIRLLAAGQVIRVPAQVSDVAVVTCVASNR 3418
QY 385 -----H-----D-----D-----L--Q----- 388
Db 3419 AGVDNKHYNLQVPAPPMNDMSMGTEBITVLKGSSTSMACITDGTAPASMAWLRDGPGLG 3478
QY 389 -----Q-----O-----H-----H-N-----KEI 394
Db 3479 DAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFILKLVLEPPHINGSSEHERI 3538
QY 395 -----YT-----I--K--W-----N----- 399
Db 3539 SVIVNNPLETCTIASGIPAPKMTWMDGRPLPQTDQVQTLGGEVLRISTAQVEDTGRYT 3598
QY 400 -----SPTG-----P--GT-----N-----N----- 408
Db 3599 CLASSPAGDDKEYLVRVHPVPPNIAGTDEPRDITVLRNQVTLCKSDAVPPPVITWLRN 3658
QY 409 -----P-----N-----ANLM-LAS--A--S--F----- 420
Db 3659 GERLOATPRVRLSGRYLQINNADLGDYANTYCVASNIAGTKTREFILTVNVPNIKGG 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433
Db 3719 POSLVILLNKSTVLECIAEGVTPPRITWRKQGAVLAGNHRVSIENGFLHIQSAHVDTT 3778
QY 434 -----C-----I--H-----TL-----T----- 439
Db 3779 GRYLCWATNAAGTDRRRIDLQVHVPPSIAPGPTNMTVIVNVQTLTACEATGIPKPSINWR 3838
QY 440 K--H-----Q-E-----P-V--Y-----S----- 447
Db 3839 KXGHLNVQNONSYRLSSGSLVIPSVDATTACTYCTVNGAGDDKRTVDLTVOVPPS 3898
QY 448 VA-----F-S-----P--D----- 453
Db 3899 IADEPTDLVTKHAPAVITCTASGVPPPSIHWTKNGIRLLPRDGYRILSSGAIEILATQ 3958
QY 454 -----GRY-----D-----KCV----- 466
Db 3959 LNHAGRYTCVARNAAGSAHRHVTLVHPEPVQPQPSSELHVILNLPILLPCEATGTPSPF 4018
QY 463 -----H-----I-W-----NT-Q--V-- 473
Db 4079 PVI5PHLKEYVIAVDKPIITLSCDAGLPPPDITWHKDGRAIVESIKQVRLSGSLQIAVF 4138
QY 474 -----C-----L-----HY----- 477

Db 4139 QPGDAGHYTCMAANVAGSSSTSKLTAVHVPPIRSTEGHYTVNENSOAILPCVADGIPTP 4198
QY 478 -----L-N-----GQ-----VL-----L-----N----- 485
Db 4199 AINWKDNVLLANLLKGYTABPYGELILENVVLEDSGFYTCVANNAAAGETHHTVSLTVHV 4258
QY 486 -----L--G--R-SIC-----L-YT-----LP-H-----H--LV-- 500
Db 4259 LPTFTLPGDVLNKGEOQLRLS-CKATGIPLPLKLTWTFNNNIIPAHFDSVNGHSELVIER 4317
QY 501 -----V--I-----P-----L-----V-----A--L----- 507
Db 4318 VSKEDSGTYVCTAENSVGFKAIGFVYVKEPPVFKGDYPSNWIETPLGGNAILNCEVKGDP 4377
QY 508 -----IE-----L-----L-VL-----K 514
Db 4378 TPTIOWNRKGVDEI5HRIRQLNGSLAIYGTVNE 4412

RESULT 5
US-10-032-189-128
; Sequence 128, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247

; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 128
; LENGTH: 5636
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3003)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3041)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3367)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; US-10-032-189-128

Query Match
Best Local Similarity 71.2%; Score 2596.5; DB 14; Length 5636;
Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

QY 1 M-S-----I-SS--D-----EV-----8
Db 1 MISWEVHTVFLFALLYSSLAQDASQSEIRAEFFEGASTLAFVFDVTGSMYDDLQVVI 60
QY 9 -----NF-IV-----12
Db 61 EGASKILETSLKRPKRLFNALFPHDPEIGPVTITDPKPFQYELRELYVQGGDCPE 120
QY 13 -----Y-----R--Y-----I-Q-----17
Db 121 MSGAIIKALIEISLPGSFIYFTDARSQYRLTHEVLQLIQKQSQVVFVLTGDCDDRTH 180
QY 18 -----E-----SG--F-----S-H-S-----A-----F 26
Db 181 IGYKVEELASTSGGVFHLDDKKQVNEVLKWEBAVQASKVHLLSTDHLEQAVNTWRIPF 240
QY 27 -----T-----FG-----I-----K 31
Db 241 DPSLKEVTVSLSGSPMIEIRNPLGLIKKGFGLHELLNIHNSAKVNVNKEPEAGMWTVK 300
QY 32 -----S--H-----I-----S--Q-----36
Db 301 TSSGRHSVRITGLSTIDFRAGFSRKTLDPKKTVSRPVQGIPTYYVLLNTSGISTPARID 360
QY 37 -----S-----NI-----N-----G-----41
Db 361 LLELLISGSSLKTI PVKYYPHRKYGIWNISDFPNPEAFELKVTGYDKDDYLFQRVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
Db 421 VFSFSIVDPAPKVTMPKTPGYILOQGIQPCSDVSLPFTLSFVRNGVTGLGVDQYLKESA 480
QY 48 -----A--L-----I-----S-----IIQ-----KG-----56
Db 481 SVSLDIKVTLSDEGFYECIAVSSAGTCRAQTFDFVSEPPVIVQPNVNTVTPGERAVLT 540
QY 57 -----L--Q-----Y-----V--E-----61
Db 541 CLIIISAVDYNLTQWRNDRVLAEPARITRLANLSLELKSVPKFNDAAGEYHCVMSSEGGSS 600
QY 62 -A-----EVSI-----NE-----68
Db 601 AASVFLTVQBPVKVTMPKQNSFTGGSEVSMCSATGYPKPKIAWTNDFVMSHRYRM 660

QY 69 --DGLTF-----D-----G-----R-----P-----I-- 78
Db 661 TSDGLFIKNAAPKADAGIYGCLANSAGTDKQNSLRYIEAPKLMVQSELVALGDITV 720
QY 79 -E--S-----L--S--LI-----DA-----V-----87
Db 721 MECKTSGIPPPQVQWFKGDLRLPSTFLIIDPLGLLKIQETQDLDAGDYTCVAINEAGR 780
QY 88 -----M-----P-DV-----VO-----T-----94
Db 781 ATGKITLDVGSPPVFIQEPADVMEIGSNVTLPCVVOGYPEPTIKWRLDNMPISRPFS 840
QY 95 -----R-----95
Db 841 VSSISQLRTGALFILNLWASDKGTTCYCEAENQFKIQSETTVTGTGLVAPLIGISPSVAN 900
QY 96 --QO-----A-----Y-----R-----D--KL-- 103
Db 901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLIQNFYITVRSDGSLHIERVOLQDGGEYTC 960
QY 104 -----A-----Q--QO-----AA-----109
Db 961 VASNVAGTNNKTVSVVHVLPFIQHQQLLSTIEGIPVTLPCKASGNPKPSVMSKKGL 1020
QY 110 ---AAA--A--A-----A--AA--A-----118
Db 1021 ISTSSAKPSAGADGSLVYVSPGEBSGEYVCTATNTAGYAKRKVQLTYYVRPRVFGDLRG 1080
QY 119 -SQ-----Q-----GS-----123
Db 1081 LSQDKPVEISVLAGEBVTLPCEVKSLLPPIITWAKETQLISPPSRHTFLPSGSMKITET 1140
QY 124 -----A-----K-----N-----126
Db 1141 RITSDSGMYLCVATNAGNTQAVKLVNHPVKIQRGPKHLKVQVGQVQVDIPCNAQGTPLP 1200
QY 127 -----GS--N-----T--A--N--G-----133
Db 1201 VITWSKGGSTMLVDGEBHVSNDPDTLSIDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260
QY 134 -----E--N-----GA-----H-----139
Db 1261 PPTVEDLEPPYNTTFOERVANORIEPPCPAKGTPKPTIKWLHNGRELTCOREPGISILEEG 1320
QY 140 T--IA-----N-----N--HT-----DM-----148
Db 1321 TLLVIASTPYDNGEYICVAVNEAGTTERKYNLKVHVPPVKDKQEQSVNVLLNQLTNL 1380
QY 149 -----M-----E-----150
Db 1381 FCEVEGTPSPIIMWYKDNVQVTTESSTIQTNNNGKILKLFRATEDAGRYSCKAINIAGTS 1440
QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160
Db 1441 QKYNIDVLVPPTIITGNFPKESVSVLNDRDVALECCQVKGTPFPDPIHFKDGGKFLFLGDPN 1500
QY 161 -----K--A-----V-----VL 165
Db 1501 VELLDRGVHLKARNRNDKGRYQCTVSNAAAGQAKDIKLIYIPPSIKGNVTDTISVL 1560
QY 166 -----RG-----H-----168
Db 1561 INSLIKLECKTRGLPMPAITWYKDGOPIMSSSQALYIDKQYLIHPAQVSDSATYTCVH 1620
QY 169 -----E--S-----E-----VF1-----C--A-----W-----177
Db 1621 ANVAGTAESFFHVVDVVPVPMIEGNLATPLNKQVIAHSLTLECNAGNPSPILTWKDGV 1680
QY 178 -----N-----P--V--SD-- 182
Db 1681 PVKANDNFRIEAGGKKLEIMSAQIEDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740
QY 183 -----LL-----ASGS-----G-----D-----S-- 191

Db 1741 TSYFIVMNNLLECHVTGSPPTIMWLKDGQLIDERDGFKILLNGRKLVIQAQVSN 1800
QY 192 -----TA-----R-----I-W 196
Db 1801 GLYRMAANTAGHKKEFEVTVHVPPTIKSSGLSERVVVKYKPVALQCIANGIPNPSITW 1860
QY 197 -----NL-----S 199
Db 1861 LKDDQPWNTAQGNLKIQSSGRVLQIAKTLLEDAGRYTCVATNAAGETQQHIQLHVHEPPS 1920
QY 200 -E-----N-----S-----SGST----- 207
Db 1921 LEDAGKMLNETVLVSNPVLQCKAAGNPVPIITWYKNCILSGSTMTFLNRQIIDIES 1980
QY 208 -Q-----L--V-----LR-H-----C-----I-- 215
Db 1981 AQISDAGIYKCAINSAGATELFYSIQVHVAPISGSNNMVAVVNNPVRLECEARGIPA 2040
QY 216 -----R-----E--G-----G-Q--D-- 221
Db 2041 PSITWLKDGSPVSSFNGLQVLSSGRILALTSTQISDTGRYTCVAVNAAGKORDIDLRV 2100
QY 222 -VPSN-----KD-----VT--S-L- 231
Db 2101 YVPPNIMBEGQNVSLISQAVELLQSDAIPPTLTWLDKGHPHLLKPKGLSISENRSVLK 2160
QY 232 -----D-----W--N--S-----EG--TLL-- 240
Db 2161 IDAQVODTGRYTCENTVAGTEKKNNVNIHWPNIIGSDBLTQLTWIEGNLSLCE 2220
QY 241 AT-----GS--Y--D-----G-----F-----A- 249
Db 2221 SSGIPPPNLWKKGSPVLTDSMGRVRIILSGGQLOISAEKSDAALYSCVASNAVAGTAK 2280
QY 250 -----R-----I-----WTXDG----- 256
Db 2281 KEYNLQVIRPTITNSGSHPTIIVTRGKSISLECEVQGIPTVTWMDGHPLIKAKGV 2340
QY 257 -----NL-AS-----T-----L-----GQH----- 265
Db 2341 EILDEGHILQKNHVSVDGRYVCVAVNAGMTDKKYDLSVHAPPSIIGNHRSPENISV 2400
QY 266 -K-----G-PI--FALK--W----- 274
Db 2401 EKNSVSLTCEASGIPLPSTTW--KDGWPVLSNVSRIILSGGRLMQTMEDACQYTC 2458
QY 275 -----N-----K-K-----GN----- 279
Db 2459 VVRNAAEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518
QY 280 -----F--I-----L-----S--AGV- 286
Db 2519 QPLQDEAHHIISSGRFLQITNVQVPHTRGYTCCLASSPAGKRSFSLNVFVSPTIAGV 2578
QY 287 -D--K--T-----T--I-----I-W----- 293
Db 2579 SDGNPDVTILNSPTSLVCEAYSYPATITWFKDGTPLSNRNRIILPGRTLOILNAQ 2638
QY 294 -D-----A-----H-----TG-----E 299
Db 2639 EDNAGRYSCVATNAEGEMIKHEVKYVIPPINKGDLWPGLSPEKVKIKVNTLTLECE 2698
QY 300 A-----K-----Q----- 303
Db 2699 AYAIPASISWYKQGPLKSDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758
QY 304 F-----P-F-----H- 307
Db 2759 FDNVIOVPPSFQKLWEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 3899 SIADBPDTFLVTKHAPAVITCTASGVPPFPIHWTKNIGIRLLPRGDCYRILSSGAIBLAT 3958

Db 2819 LTSKDKVILIPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYVRVLVPPVPIEGANS 2878
QY 316 -----WC-----SN----- 319
Db 2879 LPEEVTVLVNSKALIECLSSGSPAPRNSWQKQOQPLEDDHHKFLNGRILQILNTOITD 2938
QY 320 -----NT-----FAS--C-----S 326
Db 2939 IGRYVCVAENTAGSACKYFNLNVHVPVPSVIGPKSENLTVVVNNFISLTCEVSGFPFPPDLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKNXQPIKLTNTLIVPOGRTLQIIIRAKVSDGGEYTCIAINXAGESKKFSLTVVYPPS 3058
QY 332 --H-----V--C-----K-----L--GO-- 338
Db 3059 IKDHDSELSVNVNREGTSVLECESNAVPPVITWYKNGRMITESTHVEILADGQMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KKAESVDTGQYVCRAINVAGRDDKNFHLNVYVPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 --T-----F-----Q--G-----H-----T----- 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGSKLQIARSOHSDSGNYTCIASNMEGKAKYY 3238
QY 350 -----K-----I-W-----SM-----KQ--D-----NCV-- 384
Db 3359 TSINIECRXTGTPPOINWLKNGLPPLSSHIRLLAAGOVIRIVRAQVSDVAVYTCVASN 3418
QY 385 -----H-----Q-----D-----L--Q-- 308
Db 3419 RAGVDKNHYNQVAFAPPNMDNSMGTEETIVLKGSTSMACITDGTAPSPAMWLRDQOPLG 3478
QY 389 -----Q-----H-N-----KE 393
Db 3479 LDAHLLTVSTHGMVQLLKAETEDSGKYTCIASNEAGEVSKHFLKLEPPHINGSEHEE 3538
QY 394 I-----YT-----I--K--W----- 399
Db 3539 ISVIVNNPLELTCIASGIPAPKMTWMDKGRPLPQTDQVTLGGGEVLRISTAQVEDTRY 3598
QY 400 -----SPTG-----P--GT-----N----- 407
Db 3599 TCLASSPAGDDDKYLVLRVHVPPNIAGTDEPRDITVLRNRQVLTLECKSDAVPPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420
Db 3659 NGERLOATPRVILSGRYLOINNADLGTANTYCVASNIAGTKTREFILTNNVPPNKG 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I-- 433
Db 3719 GPQSLVILINKSTVLECIAEAGVTPTRI TWKDGAVLGNHARYSILENGFLHIQSAHVTD 3778
QY 434 -----C-----I--H-----TL-----T----- 439
Db 3779 TGRYLCMATNAAGTDRRRIDLVHVPVPSIAPGPTNMVTIVNVQTTLACBATGPKPSINW 3838
QY 440 -K--H-----Q--E-----P--V--Y----- 446
Db 3839 RKNHLLNVQDNQNSYRLLSSGLVIPSVDATYCTVNTGAGDDKRTVDLTVOVPP 3898
QY 447 SVA-----F-S-----P--D----- 453
Db 3899 SIADBPDTFLVTKHAPAVITCTASGVPPFPIHWTKNIGIRLLPRGDCYRILSSGAIBLAT 3958

```
QY 454 -----GRY-----L-----ASG--S- 461
DB 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHBPVVIQPOPSLHVILNPPILLPCEATGTPSP 4018
QY 462 F-----D-----KCV----- 466
DB 4019 FITWQEGINVTSGRNVHVLPSGGLQISRAVREDAGTYMCVAQNPAQTALGKIKLVQV 4078
QY 467 -----H-----I-W-----NT-Q----- 472
DB 4079 PPVISPHLKEYVIADVDPITLSCADGLPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 473 V-----C-----L-----HY----- 477
DB 4139 VQPGDAGHYTCMAANVAGSSSTSTKLTUVHPPRIRSTEGHYTNWNSQAILPCVADGIFT 4198
QY 478 -----L-N-----GO-----VL-----L-----N----- 485
DB 4199 PAIWKKNVLLNLLGKYTAEPYGGELLENVLEDSGFTCVANNAAGEDTHTVSLTVH 4258
QY 486 -----L-G-----R-SIC-----L-YT-----LP-H-----H-LV----- 500
DB 4259 VLPTFTLPGDVSINKGEQLRLS-CRATGIPLPKLTWTFNNIIPAHFDSVNGHSELVIE 4317
QY 501 -----V-----V-----I-PL----- 505
DB 4318 RVSKEDSGTYVTCFAENSVGFKATGFVYVKEBPVFKGDYPSNWIEPLGNAILNCEVKG 4377
QY 506 ---AL-----IE-----L-----L-VL-----K 514
DB 4378 PTFTQWRKGVDEISHRIRQLNGSLAIGTVNE 4413
```

RESULT 6

```
US-10-120-801-72
; Sequence 72, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Kekuda, Rameesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Menraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patent in Ver. 2.1
```

```
; SEQ ID NO 72
; LENGTH: 5636
; TYPE: PRT
; ORGANISM: human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3003)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3041)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3367)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; US-10-120-801-72

Query Match 71.2%; Score 2596.5; DB 14; Length 5636;
Best Local Similarity 9.7%; Pred. No. 2.7e-29;
Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

QY 1 M-S-----I-----SS--D-----EV----- 8
DB 1 MISWEVHTVTFLLYSSLAQDASPQSIRAEPEFGASTLAFVFDVTGSMYDDLQVI 60
QY 9 -----NF-LV----- 12
DB 61 EGASKILETSLKRPKRPLENFALVPHDPDEIGFVITTDPKKFQYELRELYVQGGDCPE 120
QY 13 -----Y-----R-----Y-----LQ----- 17
DB 121 MSIGAIIKIALEISLPGSFYIVFTDARKSDYRLTHEVLQIQKQSQVFLTGDCDDRTH 180
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
DB 181 IGYVVEIATSSGQVFLDKKQVNEVLKVEEAVQAKVHLLSTDHLEQAVNTWRIFP 240
QY 27 -----T-----F-----I-----K 31
DB 241 DPSLKEVTLSGSPSEMIERNPLGKIKKGFLHLLHNSAKVVNVVKEPEAGMTVK 300
QY 32 ---S-H-----I-----S--Q----- 36
DB 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSRRPVQGIPTYVLLNTSGISTPARID 360
QY 37 -----S-----N-----G----- 41
DB 361 LLELLSISGSLKTIPTVKKYPHRKYGIWNIISDFVPNEAFKLVTKYDKDYLFORVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
DB 421 VSPSSIVDPAPKVTMEKTPGYVLPQGPQPCSVDSLLPFTLSFVRNGVTGLGVQYDKESA 480
QY 48 -----A-----L-----I-----S--IIQ-----KG----- 56
DB 481 SVSLDIKAVTLSDGEFYECIAVSSAGTGRAQTFFDVSEPPVIOQVNNVTVTGERAVLT 540
QY 57 -----L--Q-----Y-----V--E----- 61
DB 541 CLIIISAVDYNLTWQRNDRVRLAEAPARITLANLSLELSKVFENDAGEYHCWVSEGGSS 600
QY 62 -A-----EVSII-----NE----- 68
DB 601 AASVFLTVQEPKVTVMKQNSFTGGSEVSIKMSATGYPKPKIAWTVNDMFIVGSHRYRM 660
QY 69 --DGTLF-----D-----G-----R-----P-----I-- 78
DB 661 TSDGTFLFIKNAAPKAGIYGLCLASNSAGTDKQNSTLRYTEAPKLMVQSELLVALGDTV 720
QY 79 -E--S-----L--S--LI-----DA-----V----- 87
DB 721 MECKTSGIPPPQVWFKGDLRLPSTFLIIDPLGLLKIQETQDLDDAGDVTCTVAINAAGR 780
```

QY 88 -----M-----P-DV-----VQ-----T-----94
Db 781 ATCKITLDVGPVPVFIQEPADVSGMEIGSNVTLPCYVQGYPEPTIKWRRLDNMFISRPFS 840
QY 95 -----R-----95
Db 841 VSSISQRTGALFILNLWASDKGTICYEABNOFGKIQSETTVTVTGLVAPLIGISPSVAN 900
QY 96 -----QQ-----A-----Y-----R-----D--KL-- 103
Db 901 VIEGQQLTLPCTLLAGNPPIPERWIKNSAMLLQNPIYTVRSOGSLHIERVOLQDGGEYTC 960
QY 104 -----A-----O--O-----AA-----109
Db 961 VASNVAGTNNKTSVVVHVHPTTQHGOQLLSTIEGIPVTLPCCKASGNPKPSVIWSKKGEL 1020
QY 110 -----AAA--A-A-----A--AA--A-----118
Db 1021 ISTSSAKFAGADGSLYVVSPEGESEGYVCTATNTAGYAKRVQLTVYVRPVFGDLRG 1080
QY 119 -SQ-----O-----O-----GS-----123
Db 1081 LSQDKPEISVLAGEEVLPCPEVKSIPPPITWAKETQLISPPSPRHTFLPSGSMKITET 1140
QY 124 -----A-----A-----K-----N-----126
Db 1141 RTSDSGMVLCVATNIAGNTQAVKLVNVHVPKIQRGPKHLKVQVGQORVDIPCNAGQTPLP 1200
QY 127 -----GE--N-----T--A-N-G-----133
Db 1201 VITWSKGGSTMLVDGHHVSNPDGTLISIDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260
QY 134 -----E--E--N-----GA-----H-----139
Db 1261 PPTVEDEPPYNTTFORVANORIEPPCPAKGTPKPTIKWLHNGRELGTREPGISILEG 1320
QY 140 T--IA-----N-----N--HT--DM-----148
Db 1321 TLLVIASVTPDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKEQVSNVLLNQLTNL 1380
QY 149 -----M-----E-----150
Db 1381 FCEVEGTPSPIIMWYKDNVQVTTESSTIQTVNNGKILKLFATPEADAGYSCRAINAGTS 1440
QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160
Db 1441 QKYFNIDVLVPPTIIGTNFPKEVSVVLRDVALLECQVKGTTPFPDIHWFKDGKPLFLGDPN 1500
QY 161 -----V-----K-A-----V-----VL 165
Db 1501 VELLDRGOVLHLKNARNDKGRYQCTVSNAGKQAKDIKLIYIPPSIKGNNVTTDISVL 1560
QY 166 -----RG-----H-----168
Db 1561 INSILKLECKTRGLPMPAITWYKDGQPIIMSSSOALYIDKQVYLHI PRAQVSDSATYCHV 1620
QY 169 -----E-S-----E-----VFI-----C--A-----W-----177
Db 1621 ANVAGTAESFHVYVVPVPIEGNLATPLNKQVIAHSLTLECNAGNPSPILTWLKGV 1680
QY 178 -----N-----P-V--SD-- 182
Db 1681 PVKANDNFRIEAGKKLEIMSAQEI DRGOYICVATSVAGEKEIKYEVVVLVPPAIEGDE 1740
QY 183 -----LL-----ASGS-----G-----D-----S-- 191
Db 1741 TSVFIWVNNLELCHVTGSPPTIMLKGDLIDERDGFILLNGRKLVIQAQVSN 1800
QY 192 -----TA-----R-----I-W 196
Db 1801 GLYRCMAANTAGDHKEFEVTVHPPTIKSSGLSERVVVVKYKPVALQCIANGIPNPSITW 1860

QY 197 -----NL-----S 199
Db 1861 LKDDQPVNTAQNLIKIQSSGRVLQIAKTLELDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920
QY 200 -E-----N-----S-----T-----SGST-----207
Db 1921 LEDAGKMLNETVLVSNPVLQCKAAGNPVPIITWYKDNCLLSGTSMTFLNRQIIDIES 1980
QY 208 -O-----L--V-----LR-H-----C--I-- 215
Db 1981 AQISDAGIYKVCAINSAGATELFYSLOVHVAPSISGSNNMVAVVNNPVPLECEARGIPA 2040
QY 216 -----R-----E--G-----G--O--D-- 221
Db 2041 PSLTWLKDGPVSSFFNGQLVLSGGRILALTSQISDTGRYTCVAVNAAGEKORDIDLRV 2100
QY 222 -VPSN-----KO-----VT--S-L-- 231
Db 2101 YVPPNIMGEEQNVSLISQAVELLQSDAIPPTLTWLKDGHPHLLKPKGLSISENSVLK 2160
QY 232 -----D-----W--N--S-----EG--TL-- 240
Db 2161 IEDAQVODTGRYTCBATNVAGTEKKNVNNVWPNIGSDELTOLTWIEGNLISLCE 2220
QY 241 AT-----GS--Y--D-----G-----F-----A-- 249
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRIILSGGRQLQISIAEKSDAALYSCVASNVAGTAK 2280
QY 250 -----R-----I-----WTXDG-----256
Db 2281 KEYNLQVYIRPTITNSGSHPTBIIVTRGKSI SLECEVOGIPPTVTVMKDGHPLIKAKGV 2340
QY 257 -----NL-AS-----T--L--GOH-----265
Db 2341 EILDEGHILQLKNHVSDDTGRYVCVAVNVAGTMDKKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K-----G-PI--FALK--W-----274
Db 2401 EKNSVSLTCEASGIPLPSTTW--KDGWPSVLSNSVRIILSGGRMLRMLQTTMEDAQYTC 2458
QY 275 -----N-----K-K-----GN-----279
Db 2459 VVRNAAGEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPPIETHKDG 2518
QY 280 -----F--I-----L--S--AGV-- 286
Db 2519 QPLQDEAHIIISGGRFLQITNVQVPHTRYTCCLASSPAGHKSRSPSLNVFVSPITAGVG 2578
QY 287 -D--K--T-----T--I-----I-W-----293
Db 2579 SDGNFEDVTILNSPTSLVCEAVSYPPATITWFKDGTPLSNRNRIILPGRQLQILNAQ 2638
QY 294 -D-----A--H-----TG-----E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKYIPIINKGDLWGLSPKEVKIKVANTLTLECE 2698
QY 300 A-----K-----O--Q-----303
Db 2699 AYAI PASLSWYKDGQPLKSDDDHVNIAANGHTLQI KEAQISDTGRYTCVASNIAGEDLD 2758
QY 304 F-----P-F-----H-- 307
Db 2759 FDNVTOVPSFQKLWEIGNMLDTGRNGEAKDVIINNPI SLYCETNAAPPTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGDSLQYDVRVLVPPPIEGANS 2878
QY 316 -----WQ-----SN-----319
Db 2879 LPEEVTVLNKSALIECLSSGSPAPRNSWQDQFLLEDHDKFLSNGRILQILNTQITD 2938
QY 320 -----NT-----FAS--C--S 326

Db 2939 IGRYVCAENTAGSAKYNFLNHHVPPSVIGPKSENLTVVNNFISLTCEVSGFPDPDL 2998
QY 327 - - - - -T- - - - -D- - - - -MCI- - - - - 331
Db 2999 WLKNXQPIKLTNTLTPVPGRTLIQIRAKVSDGGEVTCIAINXAGBSKKKFSLTVVVPPS 3058
QY 332 - - - - -H- - - - -V- - - - -C- - - - -K- - - - -L- - - - -GO- - - - - 338
Db 3059 IKDHDSELSVNVNREGTSVLECSNAVPPVITWYKNGRMITESTHVEILLADGQMLHI 3118
QY 339 - - - - -D- - - - -R- - - - -P- - - - -IK- - - - - 343
Db 3119 KKAESDGTQGVCRAINVAGRDKNFHLNVVPPSVIEGPEREVIVTISNPVTLTCDATG 3178
QY 344 - - - - -T- - - - -F- - - - -Q- - - - -G- - - - -H- - - - -T- - - - - 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGSKLQIARSQSDSGNYTCIASNMKGAKQYV 3238
QY 350 - - - - -K- - - - -I- - - - -W- - - - -SM- - - - -IKW- - - - -D- - - - -P- - - - -T- - - - - 360
Db 3239 FLSIQPPSVAGAEIPSDVSLGENVELVCNANGIPTPLIOWLKDKGFIASGETERIRV 3298
QY 361 - - - - -G- - - - -NLL- - - - -A- - - - -S- - - - -C- - - - -S- - - - -D- - - - -D- - - - -MTL- - - - - 373
Db 3299 SANGSTLNIYGALTSDTGKTKCVATNPAGEEDRIFNLNVVVTPIRGNKDEAKLMTYVD 3358
QY 374 - - - - -K- - - - -I- - - - -W- - - - -SM- - - - -KO- - - - -D- - - - -NCV- - - - - 384
Db 3359 TSINIECRXTGTPPPQINLKNGLPLPLSHIRLLAAGQVIRIVRAQVSDVAVYTCASN 3418
QY 385 - - - - -H- - - - - 388
Db 3419 RAGVDNKHYNLQVAPPNMDNSMGTEITVLKGSSTSMACITDGTAPSNWAWLRDQPLG 3478
QY 389 - - - - -DAHLTVSTHGMVQLLKAEDESGKYTCIASNEAGEVSKHFLKVLPEPHNGSEHEE 3538
QY 394 I- - - - -YT- - - - -I- - - - -K- - - - -W- - - - - 399
Db 3539 ISVIVNNPLELTCTASGIPAPKMTWKMKGRLPQTDQVOTLGGGEVLRISTAQVEDTRY 3598
QY 400 - - - - -SPTG- - - - -P- - - - -CT- - - - -N- - - - - 407
Db 3599 TCLASSPAGDDKEXLVRVHVHPNIAGTDEPRDITVLRNQVTLCKSDAVPPPVITWLR 3658
QY 408 N- - - - -P- - - - -N- - - - -ANLM- - - - -LAS- - - - -A- - - - -S- - - - -F- - - - - 420
Db 3659 NGERLOATPRVILSGRYLIQINNADLGDYANTCVASNIAGKTREFILTVNVPPNKG 3718
QY 421 - - - - -DSTV- - - - -RL- - - - -W- - - - -D- - - - -V- - - - -DR- - - - -G- - - - -I- - - - - 433
Db 3719 GPOSVLINKSTVLECIAGEVTPTRITWRKQGAVLGNHARYSILENGFLHIQSAHVTD 3778
QY 434 - - - - -C- - - - -I- - - - -H- - - - - 439
Db 3779 TGRYLCMATNAAGTDRRRIDLVHVPPSVIAPGPTNMTVIVNVQTLTACEATGPKPSINW 3838
QY 440 - - - - -K- - - - -H- - - - -Q- - - - -E- - - - -P- - - - -V- - - - -Y- - - - - 446
Db 3839 RKNGHLLNVNDQNSYELLSSGSIIVISPSVDDTATVECTVNGAGDKETVLTQVVPF 3898
QY 447 SVA- - - - - 453
Db 3899 SIADEPTDLVTKHAPAVITCTASGVFPFSIHMTKNGIRLLPRGDGVRILSSGAIEILAT 3958
QY 454 - - - - -GRY- - - - - 461
Db 3959 QLNHAGRYTCVARNAGSAHRHVTLVHVEPPVQPPQPSSEHLVILNPNILLPCEATGTPSP 4018
QY 462 F- - - - - 466

Db 4019 FITWQKEGINVNTSGRHNHVLPSGGLQISRAVRVEDAGTYMCVAQNPAQTALGKIKLNVQV 4078
QY 467 - - - - -H- - - - -I- - - - -W- - - - - 472
Db 4079 PPVISPHLKEYVIADVDPITLSCEADGLPPDPITWHKQGRAIVESIRQRLVSSGLQIAF 4138
QY 473 V- - - - -C- - - - -L- - - - -HY- - - - - 477
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIIRSTEGHYTVNENSOAILPCVADGPT 4198
QY 478 - - - - -L- - - - -N- - - - -GO- - - - -VL- - - - -L- - - - -N- - - - - 485
Db 4199 PAINWKONVLLANLILGKYTAEPYGBELIENVVLEDSGFYTCVANNAGEDTHVSLTVH 4258
QY 486 - - - - -L- - - - -G- - - - -R- - - - -SIC- - - - -L- - - - -YT- - - - -LP- - - - -H- - - - -LV- - - - - 500
Db 4259 VLPTTFELFGDVLNKGEOQLRLS-CKATGIPLPKLTWTNNIIPAHFDSVNGHSELVIE 4317
QY 501 - - - - -AL- - - - -IE- - - - -L- - - - -L- - - - -VL- - - - -K- - - - - 514
Db 4318 RVSKEDSGTYVCTAENSUGFVKATGFVYVKEPPVFKGDPSPNWIPLGGNAILNCEVKGD 4377
QY 506 - - - - -AL- - - - -IE- - - - -L- - - - -L- - - - -VL- - - - -K- - - - - 514
Db 4378 PPTIQMNRKGVDDIEISHRIROLGNGSLAIYGTVNE 4413

RESULT 7
US-10-023-634-93
; Sequence 93, Application US/10023634
; Publication No. US20030236389A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Ballinger, Robert A
; APPLICANT: Guo, Xiaojia
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Li, Li
; APPLICANT: Ellerman, Karen
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Burgess, Catherine E
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gangolli, Bsha A
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Gunther, Erik
; APPLICANT: Smithson, Glenda
; APPLICANT: Millet, Isabelle
; APPLICANT: Gerlach, Valerie
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-221
; CURRENT APPLICATION NUMBER: US/10/023,634
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: 60/256,025
; PRIOR FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: 60/265,163
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/272,929
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/274,864
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/276,688
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,880
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/286,409
; PRIOR FILING DATE: 2001-04-25

; PRIOR APPLICATION NUMBER: 60/309,246
 ; PRIOR FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: 60/315,600
 ; PRIOR FILING DATE: 2001-08-29
 ; NUMBER OF SEQ ID NOS: 132
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 93
 ; LENGTH: 5636
 ; TYPE: PR1
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (3003)
 ; OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
 ; OTHER INFORMATION: specification.
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (3367)
 ; OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
 ; OTHER INFORMATION: specification.
 ; US-10-023-634-93

 Query Match 71.2%; Score 2596.5; DB 14; Length 5636;
 Best Local Similarity 9.7%; Pred. No. 2.7e-29;
 Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

 QY 1 M-S-----I-----SS--D-----EV-----8
 DB 1 MISWEVHTVFLFALLYSSLAQASQSEIRAEPEPEGASTLAFVFDVTGSMYDVLQVVI 60
 QY 9 -----NF-LV-----12
 DB 61 EGASKILETSKRPRPLFNALVPFHDPEIGPVTTTDPKQYELRELYVGGDCPE 120
 QY 13 -----Y-----R--Y-----IQ-----17
 DB 121 MSIGAIALEISLPGSFYVFTDARSQYRLTHEVLQIQKQSQVVFVLTGDCDDRTH 180
 QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
 DB 181 IGKVEEETASTSGGVFHLDDKQVNEVLKWVEEAVQASKVHLLSTDHLEQAVNTWRIEP 240
 QY 27 -----T-----FG-----I-----K 31
 DB 241 DPLKEVTVSLGSPSPMIEIRNPLGLIKKGFGLHELLNIHNSAKVNVNKEPEAGMWTVK 300
 QY 32 -----S--H-----I-----S--Q-----36
 DB 301 TSSGRHVRITGLTIDFRAGFSRKPTLDFKKTVSRRPVQGIPTYVLLNTSGISTPARID 360
 QY 37 -----S-----NI-----N-----G-----41
 DB 361 LLELLSISGSSSLTIPVKYYPHRKPYGIWNIISDFVPNEAFFLKVTGYDKDYLFORVSS 420
 QY 42 -----A-----L-----VP-----P-----A 47
 DB 421 VSFSSIVDPKVTMPEKTPGYLQPGQIPCSVDLSLPTLSPVRNGVTGLGVQYLKESA 480
 QY 48 -----A-----L-----I-----S-----IIQ-----KG-----56
 DB 481 SVSLDIKVTLSDEGYECIAVSSAGTGRAQTFDFVSEPPVVIQVNNVTVPGERAVLT 540
 QY 57 -----L--Q-----Y-----V--E-----61
 DB 541 CLIIISVDYNTWQRNDRVRLAEPARIRTLANLSLELKSVPFNDAAGEYCHWVSEGGSS 600
 QY 62 -A-----EVSI-----NE-----68

Db 601 AASVFLTVQEPKPVTVMPKNSQFTGGSEVSMCSATGYPKPKIAWTVNDMFIVGSHRYRM 660
 QY 69 --DGLTF-----D-----G-----R-----P-----I--78
 Db 661 TSDGTLFIKNAAPKDAIGYGLASNSAGTDKQNSTURYIEAPKLMVVOSELLVALGDIIV 720
 QY 79 -E--S-----L--S--LI-----VO-----T-----87
 Db 721 MECKTSGIPPPQVWKFGDLELRPSTFLIDPLGLLKIQETQDLDAGDYTCVAINAGR 780
 QY 88 -----M-----P-DV-----94
 Db 781 ATGKITLDVGGPPVFIQEPADVSMGICSNVTLPCTYVQGYPEPTIKWRRLDNMPISRPFS 840
 QY 95 -----R-----95
 Db 841 VSSISQIRTGALFILNLWASDKGTICEAENQPKIQSETTVTVTGLVAPLIGISPVSAN 900
 QY 96 -----QQ-----A-----Y-----R--D--KL--103
 Db 901 VIEGQQLTLPCTLLAGNPIPERRWIKNSAMLLQNPIYTVRSQGLHIERVOLQDGEYTC 960
 QY 104 -----A-----Q--Q-----AA-----109
 Db 961 VASNAGTNNKTTVVVHVLPFTIQHGQILSTIEGIPVTLPCCKASGNPKPSVIWSKKGEL 1020
 QY 110 ---AAA---A-A-----A--AA--A-----118
 Db 1021 ISTSSAKFAGADGSLYVVSPEGESEGYVCTATNTAGAKRVQLTYYVVRPVGDLRG 1080
 QY 119 -SQ-----A-----K-----N-----126
 Db 1081 LSQDKPEISVLAGEEVLTPCEVKSLPPPIITWAKETQLISPPSPRHTFLPSSMKITET 1140
 QY 124 -----A-----A-----K-----N-----126
 Db 1141 RTSDSGMVLCVATNIAGNVTAQVKNLVHVPKIQRGPKHKVQGVQKQVPCNAQGTPLP 1200
 QY 127 -----GE--N-----T--A--N--G-----133
 Db 1201 VITWSKGSMTLVGDEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDEITLHVQE 1260
 QY 134 -----E--N--N-----GA-----H-----139
 Db 1261 PPTVEDLEPPYNTTFOERVANQRIEPPCPAKGTPKPTIKWLHNGRELTGREPGISILEEG 1320
 QY 140 T--TA-----N-----N--HT--DM-----148
 Db 1321 TLLVIASVTPDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKQVSNVSVLLNQLTNL 1380
 QY 149 -----M-----E-----150
 Db 1381 FCEVEGTPSPIIMYKNDNVQVTSSTIOTVNNKGILKLFATPDEADGRYSCKAINAGTS 1440
 QY 151 -----VD-----G-----DV--E-----I-----P--PN 160
 Db 1441 QKYNFIDVLVPTIIGTNPKEVSVVLRDVALEQCVKGTPTPDHIFWPKDGLFLGDPN 1500
 QY 161 -----K-A-----V-----VL 165
 Db 1501 VELLDRGQVLHLKNARRNDKGRYQCTVSNAGKQAKDIKLTIVIPPSIKGNVTTDISVL 1560
 QY 166 -----RG-----H--168
 Db 1561 INSLIKLECKTRGLPMPAITWKDQPIIMSSSQALYIDKQYLHIHPRQVSDSATYCHV 1620
 QY 169 -----E--S-----E-----VFI-----C--A-----W-----177
 Db 1621 ANVAGTAEKSFHVTVVPPMIEGNLATPLNKQVIAHSLTLECNAGNPSPILTWLDKGV 1680
 QY 178 -----N-----P--V--SD--182
 Db 1681 PVKANDNPFIEAGKKLEIMSAQIDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740

QY 183 -----LL-----ASGS-----G-----D-----S-- 191
Db 1741 TSYFIYVWNNLELCHVTGSPPTIMWLKQGLIDRDERGFKILLNGRKLVIQAQVNT 1800
QY 192 -----TA-----R-----I-W 196
Db 1801 GLYRCMAANTAGDKKEFEVTHVPTIKSSGLSERVVVYKFPVALQCIANGIPNPSITW 1860
QY 197 -----NL-----S-----S 199
Db 1861 LKDDQPNTAQGNLKIOSSGRVLIQIAKTLLEDAGRVCVATNAAGETQOQHVLHVPEPS 1920
QY 200 -E-----N-----S-----SGST----- 207
Db 1921 LEDAGKMLNETVLVSNPVLQLECKAAGNPVPTIWKYKDNCLLGGSTSWTFLNRQIIDISS 1980
QY 208 -Q-----L--V-----LR-H-----C-----I-- 215
Db 1981 AQISDAGIKVCVAINSAGATELFYSLOVHVAPSISGSNNMVAVVNNPVRLCEARGIPA 2040
QY 216 -----R-----E-----G-----Q-----D----- 221
Db 2041 PSLTWLKDGPVSSFSFNGVLQVLSGGRILALTSQISDTGRYTCVAVNAAGEKQORDILRV 2100
QY 222 -VPSN-----KD-----VT---S-L- 231
Db 2101 YVPPNMGEEQNVSVLISOAVELLCSQDAIPPTTLTWLKDGHPLKPKGLSISENSVLK 2160
QY 232 -----D-----W--N--S-----EG--TLL-- 240
Db 2161 IEDAQVDTGRYTCBATNVAGKTEKQNVNINWPNIGSDELTOITWIEGNLISLCE 2220
QY 241 AT-----GS--Y--D-----G-----F-----A- 249
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVILSGGRQLQISIAEKSDAALXSCVASNAGTAK 2280
QY 250 -----R-----I-----WTQDG----- 256
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRKGSISLCEVQGPPTVTWMDGHPLIKAKGV 2340
QY 257 -----NL-AS-----T-----L-----GOH----- 265
Db 2341 EILDEGHILQKNIHVSQDGRYCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPEINISVV 2400
QY 266 -K-----G-PI-----FALK--W----- 274
Db 2401 EKNSVSLTCEASGILPLPSTTFW--KDGWPVSLNSVRILSGGRMLRMLMOTTMEDAQYTC 2458
QY 275 -----N-----K-K-----GN----- 279
Db 2459 VVRNAGEERKIFGLSVLPVPHIVGENTLEBDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518
QY 280 -----F--I-----I-W-----L--S--AGV-- 286
Db 2519 OPLQDEAHIIISGGFLOITNVQVPHTRGYTCLASSPAGKSRSPSLNVFVSPTIAGVG 2578
QY 287 -D--K--T-----T-----I----- 293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTPLSERNRILRILPGRTLOILNAQ 2638
QY 294 -D-----A-----H-----TG-----E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYKVYIPPIINKDLWGLSPKXVKIKVNNLTLECE 2698
QY 300 A-----K-----Q-----Q----- 303
Db 2699 AYAIPSASLSWYKDGQPLKDDHVNIAANGHTLOIKEAQISDGTGRYTCVASNAGEDLD 2758
QY 304 F-----P-F-----H- 307
Db 2759 FVNIQVPPSFQKLEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPPTLTWYKDGHP 2818

QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTSSDKVILPGRVLOIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPIIEGANS 2878
QY 316 -----WQ-----SN----- 319
Db 2879 LPEEVTLVNKSALIECLSSGSPAPRNSWQDQPLEDDHKKFLSNGRILQILNTQITD 2938
QY 320 -----NT-----FAS--C-----S 326
Db 2939 IGRYVCVAENTAGSACKYFNLVNHPVPSVIGPKSENLTVVVNNFISLTCEVSGFPPDLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKNQPIKLTNTLIVPGRTLQIIRAKVSDGGEYTCIAINXAGESKKKFSITVYVPPS 3058
QY 332 --H-----V--C-----K-----L--GO-- 338
Db 3059 IKDHDSESLSVVNVREGTSVLECESNAVPPVITWYKNGRMITESTHVEIILADGQMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KKAESDGTQYVCRAINVAGRDDKNPHLVNYPVPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 -----T-----F-----Q-----G-----H-----T----- 349
Db 3179 IPPPTIAMLKNYKRIENSLSVRILSGSKLOIARSQSDSGNYTCIASNMEGKAQYV 3238
QY 350 -----G--NLL-A--S-----N-E-V-NA-----IKW--D--P-----T----- 360
Db 3239 FLSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIOWLKDGPDIASGETERIRV 3298
QY 361 -----G--NLL-A--S-----C--S-----D-----D-----MTL-- 373
Db 3299 SANGSTNIYGALTSDTGKYTCVATNPAGEDRIFNLNVYVTTIRGNKDEAKMLTYVD 3358
QY 374 -----K-----I-W-----SM-----KQ--D--D-----NCV-- 384
Db 3359 TSINIECRXTGTPPQINLKNGLPLPLSSHILLAAQOVIRIVRAQVSDVAVYTCASN 3418
QY 385 -----H-----D-----L--Q----- 388
Db 3419 RAGVDKNHYLQVAPPNNMDSMGTEETVLKGSSTSMACITDGTTPAPSMWLRDQPLG 3478
QY 389 -----Q-----H-N-----H-N-----KE 393
Db 3479 LDAHLTVSTHGMVLQLLKAETDSGKYTCIASNEAGEVSKHFIKLVLEPPHINGSEHEE 3538
QY 394 I-----YT-----I--K--W-----N----- 399
Db 3539 ISVIVNPLELTICIASGIPAPKMTWMDGRPLPQTDQVTLGGGEVLRISTAQVEDTRY 3598
QY 400 -----SPTG-----P-----GT-----N----- 407
Db 3599 TCLASSPAGDDKEYLVRVHVPPNIAGTDEPRDITVLRNRQVTLCKSDAVPPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420
Db 3659 NGERLOATPRVILSGRYLOINNADLGTANTYTCVASNIAGTKTREFILTVMVPPNKG 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433
Db 3719 GPQSLVILLNKSTVLECIAEGVPTPRIWTKDGAVALGNHARYSILENGLFHQSAHVT 3778
QY 434 -----C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCMATNAAGTDRRIDLQVHVPPSIAPGPTNMTVIVNVQTTILACEATGPKPSINW 3838
QY 440 -K--H-----Q-E-----P-V-----Y----- 446
Db 3839 RKNGHLLNVQDQNSYRLLSSGLVSIISPSVDDTATYECTVTNGAGDDKRTVLTVOVPP 3898
QY 447 SVA-----F-S-----P--D----- 453

Db 1141 RTSDSGWLCVATNIAGNVTAQVAKLVNHHVPPKIQRGKHLKVQVQGRVDIPCNQAQGTPLP 1200
QY 127 -----GE-----N-----T--A-N-G----- 133
Db 1201 VITWSKGGSTMLVDGHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDEBITLHVQB 1260
QY 134 -----E-----E-N-N-----GA-----H----- 139
Db 1261 PPTVELEPPYNTTFQERVANQRIEPPCPAKGTPKPTIKWLHNGRELGTREPGISILEEG 1320
QY 140 T-----IA-----N-----N-----HT-----DM----- 148
Db 1321 TLLVIASTVPDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDEQVSNVSLNLQTLN 1380
QY 149 -----M-----N-----E----- 150
Db 1381 FCEVEGTPSPPIIMWYKDNVQVTSSTIQTVNNKILKLFATPEDAGRYSCKAINIAGTS 1440
QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160
Db 1441 QKYNIDVLVPPTIIGTNPPEKVSUVRNDRDALECCQVKGTPFPDIIHWFKDGKFLGDPN 1500
QY 161 -----K-A-----K-A-----V-----VL 165
Db 1501 VELLDRGOVLHLKNARENDKGRYQCTVSNAAGQAQDKLTIYI PPSIKGNVTTDISVL 1560
QY 166 -----RG-----H----- 168
Db 1561 INSLIKLECKTGLPMPAITWYKDGQPIIMSSQALYIDKQYLHI PRAQVSDSATYTCV 1620
QY 169 -----E-S-----E-----VFI-----C-A-----W----- 177
Db 1621 ANVAGTAKSFHVDVVPVPMIEGNLATPLNKQVIAHSLTLECNAGNPSPIITWLDGV 1680
QY 178 -----N-----P-V--SD-- 182
Db 1681 PVKANDFRIEAGGKLEIMSAQEI DRGOYICVATSVAGEKEIKEYVDVLVPPAIEGGDE 1740
QY 183 -----LL-----ASGS-----G-----D-----S-- 191
Db 1741 TSYFIVMNNLELDCHTVGTSGPPPTIMWLKDGQLIDERDGFKILLNGRKLVIQAQVSN 1800
QY 192 -----TA-----R-----I-W 196
Db 1801 GLYRCHMAANTAGDHKEFEVTVHVPPTIKSSGLSERVVVVKYKVALQCIANGIPNSITW 1860
QY 197 -----NI-----S 199
Db 1861 LKDDQPNTAQGNLKIQSSGRVLQIAKTLLDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920
QY 200 E-----N-----S-----T-----SGST----- 207
Db 1921 LEDAGKMLNETVLSNPFVLECKAAGNPVPIWYKDNCLLGSSTMTFLNRQIIDIES 1980
QY 208 Q-----L--V-----LH-H-----C-----I-- 215
Db 1981 AQISDAGIYKVAINSAGATELFVSLQVHVAPISGSNNMVAVVNPNVPLECEARGIPA 2040
QY 216 -----R-----E-----G-----G-Q--D-- 221
Db 2041 PSLTWLKDGPVSPFSNGLQVLSGRILALTSQISDTGRYTCVAVNAAGEKQRDIDLVR 2100
QY 222 VPSN-----KD-----VT-----S-L- 231
Db 2101 YVPPNIMGEQNVSVLISQAVELLCSDAIPPTPLTWLKDGHPLLLKPKGLSISENSVLK 2160
QY 232 -----D-----W-----N-----S-----EG--TLL-- 240
Db 2161 IEDAQVODTGRYTCEATNVAGTEKKNNVNIWPPNIGSGDELTLQTLVIEGNLISLCE 2220
QY 241 AT-----GS--Y-D-----G-----F-----A- 249
;

Db 2221 SSGIPPPNLIWKKGSPVLTDMSGRVRIILSGRQLQISIAEKSDAALYSCVASNVAGTAK 2280
QY 250 -----R-----I-----WTKG----- 256
Db 2281 KEYNLQYIRPITITNSGSHPTBIIIVTRGKSIISLECEVQIQIPPTVTWMDGHPHAKGV 2340
QY 257 -----NL-AS-----T-----L-----GQH----- 265
Db 2341 EILDEGHILQAKNIHSDTGRYCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K-----G-PI--W-----FALK--W----- 274
Db 2401 EKNSVLTCEASGIPLPSTTF--KDGWPSVLSNVRILSGRMLRMOTTMEDAQYTC 2458
QY 275 -----N-----K-K-----GN----- 279
Db 2459 VVRNAGEBERKIFGLSVLPVPHVGTENTLEDVKVKEKQSVTLTCEVTGNPVBETWHKDG 2518
QY 280 -----F-I-----L-----S--AGV-- 286
Db 2519 QPLQEDEAHHIISGGFRLOITNVQVPHGTGRYTCLASSPAGHKSRSPSLNVFVSPTIAGVG 2578
QY 287 -D---K--T-----T--I--I-W----- 293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLSNRNRIRILPGRTLQILNAQ 2638
QY 294 -D-----A-----H-----TG-----E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVYIPIINKGLWPGLSPEKVKIKVNTLTLECE 2698
QY 300 A-----K-----Q-----Q----- 303
Db 2699 AVAIPASLSWYKDGQPLKSDDDHVNIAANGHTLOIKEAQISDTGRYTCVASNIAGEDEL 2758
QY 304 F-----P-F-----H- 307
Db 2759 FDNVIOVPPSFQKWEIGNMLDTRNGEAKDVIIINNPISLYCEETNAAPPTLTWYKDGH 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTSSDKVILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878
QY 316 -----WO-----SN----- 319
Db 2879 LPBEVTVLNKSALIECLSSGPAPRNSWQKQOPLLEDHKKFLSNGRILQILNTQITD 2938
QY 320 -----NT-----FAS--C-----S 326
Db 2939 IGRYVCVAENTAGSAKKYFNLVHVPVPSVIGPKSENLTVVNNFISLTCEVSGFPDPLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKXQPIKLTNTLIVPGRTLQIIRAKVSDGGEYTCIATNAXGSKKKFSLTVVYVPPS 3058
QY 332 --H-----V--C-----K-----L-GO-- 338
Db 3059 IKDHDSELSVNVNREGTSVLECESNAVPPVITWYKNGRMITESTHVEILADGOMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KKAESVDTGOYVCRAINVAGRDDKNFHLNVYVPPSIEGPEREVITISNPVTLTCDATG 3178
QY 344 --T-----F-----Q--G-----H-----T----- 349
Db 3179 IPPTTIAWLKQYKRIENSLSLEVRILSGSKLQIARSQSDSGNYTCIASNMEGKAKYY 3238
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360
Db 3239 FLISIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIOWLKDGPKTASGETERIV 3298
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL-- 373
Db 3299 SANGSTINIYALTSDTKYTCVATNPAGEEDRIFNLNVYVPTIRGNKDEAKLMTYVD 3358

```
QY 374 ---K---I-W---SM---KQ-D---NCV--- 384
Db 3359 TSINIECRXTGTPPPQINLWKNGLPLSLSHIRLAAGQVIRIVRAQVSDVAVYTCVASN 3418
QY 385 ---H---D---L-Q--- 388
Db 3419 RAGVDNKNHYLQVAFAPNDMSMGTEITVLKSGSTSMACITDGTAPASMAWLRDQPLG 3478
QY 389 ---Q---H-N---KE 393
Db 3479 LDAHLYTSHGMVQLLKARTDSGKTCIASNEAGEVSKHFILKVLPPHINGSSEHEE 3538
QY 394 I---YT---I-K-W--- 399
Db 3539 ISIVVNNPLELTCIASGIPAKMTWMDGRPLQTDQVQLTGGGEVLRISTAQVEDTGRY 3598
QY 400 ---SPTG---P-GT---N--- 407
Db 3599 TCLASSPAGDDDKKEYLVRVHVHPNIAGTDEPRDITVLNRNQVLTCECKSDAVPPPVITWLR 3658
QY 408 N---P---N---ANLM-LAS-A-S-F--- 420
Db 3659 NGERLOATPRVILSGRYLQINNADLGTANTYTCVASNTAGTKTREFILTVNVPNIKG 3718
QY 421 ---DSTV---RL-W-D-V---DR---G-I--- 433
Db 3719 GPQSLVILLNKSTVLECIAGVTPRITWRKQGANVLGNHARYSILENGELHIOASHVTD 3778
QY 434 ---C---I---H---TL---T--- 439
Db 3779 TGRYLCMATNAAGTDRRRIDLVHVPPSIAPGPTMTVINVTTLACEATGPKPSINW 3838
QY 440 -K-H---Q-E---P-V---Y--- 446
Db 3839 RKNGLLNVDQNSYRLSSGLVLIISPSVDDTATYECTVTNGAGDKRTVDLTQVVP 3898
QY 447 SVA---F-S---P-D--- 453
Db 3899 SIADPTDFLVTKHAPAVITCTASGVFPFSIHWTKNIGIRLLPRGDGYRILSSGAIBLAT 3958
QY 454 ---GRY---L---ASG-S- 461
Db 3959 QLNHAGRYTCVARNAAGSAHRVTLHVHEPPVIOPOPSSELHVLNPNILLPCEATGTPSP 4018
QY 462 F---D---KCV--- 466
Db 4019 FITWQKEGINVNTSGRNVAVLPSSGLQISRAVEDAGTYMCVAQNPAQTALGKIKLVQV 4078
QY 467 ---H---I-W---NT-Q--- 472
Db 4079 PPVISPHLKEYVIADKPTLSCADGLPPDPDITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 473 V---C---L---HY--- 477
Db 4139 VQPGDAGHYTCMAANVAGSSSTKLTVHVPPRIRSTEGHYTNENSQAILPCVADGPT 4198
QY 478 ---L-N---GO---VL---L---N--- 485
Db 4199 PAINWKKDNLVLANLLGKYTAEPYGBILENVVLEDSGFYTCVANNAAAGDTHVSLTVH 4258
QY 486 ---L-G---R-SIC---L-YT---LP-H---H-LV--- 500
Db 4259 VLPTFTELPGDVSINKGEQURLS-CRATGIPLPKLTWTFNNNIIPAHFDVSNGHSELVIE 4317
QY 501 ---V---I-PL---V--- 505
Db 4318 RVSKEDSGTYCTAENSVGPKATGFVYVKEPFPKGDYPSNWIEPLGGAAILNCEVKGD 4377
QY 506 ---AL---IE---L---L-VL---K 514
Db 4378 PTPTIQNRKGVDEISHRIRQLNGSLATYGTWNE 4413
```

```
RESULT 9
US-10-464-368-69
; Sequence 69, Application US/10464368
; Publication No. US20040023356A1
; GENERAL INFORMATION:
; APPLICANT: Krumlauf, Robb
; APPLICANT: Ellies, Debra
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
; FILE REFERENCE: 40716-IP-017
; CURRENT APPLICATION NUMBER: US/10/464,368
; CURRENT FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/388,970
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 69
; LENGTH: 4599
; TYPE: PR1
; ORGANISM: MOUSE
US-10-464-368-69
```

Query Match 71.0%; Score 2589.5; DB 15; Length 4599;
Best Local Similarity 9.3%; Pred. No. 1.3e-29;
Matches 428; Conservative 65; Mismatches 14; Indels 4095; Gaps 370;

```
QY 1 MS-----I-----S----- 4
Db 1 MSOLLAILTLGLLPNAEVLIVGANODQHLCDPGBFLCHDHVTCVSQSLCDGDDPCPD 60
QY 5 -SD-----EV-----N-----F----- 10
Db 61 QSDSELDTCPEEVEIKPLNHIACHGSSACVHLSKLCNGVVDPCDFDEGGHCQELLPSC 120
QY 11 --L---V---Y---R---Y-L-Q--- 17
Db 121 QQLNCFKCMVRNATRCYCEGFEVAEDGRCKDQDECSIYGCISQCTCKNTYGSVACSC 180
QY 18 -ESGF---S---H---SAF-T---F---G-K---S---H--- 33
Db 181 VE-GYIMQSDNRCKVKHEPTDKAPMLLISSLETIELFYINGSKMTTLSSANRNEIHTLD 239
QY 34 -----I-S-QS-----NI--- 39
Db 240 FYISEEMICWIESRESSNQLKCGKITKAGLTDRILNSLQSFQNVQMAFDMLTRNIYF 299
QY 40 -----NG-----A-L---VP--- 45
Db 300 VDHVSDRIFCVNFNGSVCVTLIESELHNPKAATAADPIAGKLFFTDYCNVPKIERCDLDGM 359
QY 46 -----P-A-AL-----SIIQKGLQ---Y- 59
Db 360 NTRIVYKAEQPSALALDLVNRLLVYVDLYLDYGVVDYQGNRHTIVQ-GRQVKHLYG 418
QY 60 --V-E---AEVS---IN-----E- 68
Db 419 ITVFEDYLYA-TSSDNFNIRNFRNGTDIHSIIKMSARGIRTYQKRTQPTVRSACEV 477
QY 69 D-----G-----TL---F---DGR--- 76
Db 478 DAYGMPGCCSHICLLSSSYKTRTCRCRTGTFNMGSDGRSKCRPKNELFLFYKGRPGIVRG 537
QY 77 -----PIE---SL-----S-LI----- 84
Db 538 MDLNTKIADECMPIENLPRALDFAEANYIYFADTTSFLGRQKIDGTERETILKDD 597
QY 85 -D---AV-----M-----P--- 89
Db 598 LUNVEGIAVDWIGNNLYWTNDGHRKTIINVARIABLEKASQSRKTLLEGSGHPRAIWVDPNG 657
QY 90 -----D-V---V-Q--- 93
Db 658 WMYWTDWKEDKIDDSVGRIEKAMMDGVNRQVFTVSKMLWPNGLTLDPHTSTLYWCDAIYD 717
```

QY 94 -----T-R-----Q----- 96
Db 718 HIBKVLNGTHRKVVISGKELNHPFGLSHHGNVFWFTDYNGSIFQDLMTNEVTLRHE 777
QY 97 -----QAY--R--D-K-----LA----- 104
Db 778 RAPLGLQIYDPRKQGDNNCRINNGCGGTLCLAIAPAGRYCACADNOLLDENGTTCFNP 837
QY 105 -----Q-----Q-----Q-----Q----- 107
Db 838 EEIRPHICKGEFRCKNKHCIQARWKCDGDDCLDGSDEDSVTCFNHSCPDQFKQNNR 897
QY 108 -----A-----AA--AA-----AA-----AA-116
Db 898 CIPKRWLGDGANDCGNEDESNOTCTARTQAOQFSCGNRCIPTAWLDCDREDDCGDQTD 957
QY 117 -AAS-----Q-----Q-----Q-----Q----- 121
Db 958 EVASCEPPTCEPLTQFICKSGRCISNKHWCDDTDDCCDRSDEVCVHSCLDQFRCSSGR 1017
QY 122 ---GS-A---KN-G---E---N-T---A-----N-----GE-134
Db 1018 CIPGHWACDNDGDFDSDETHINCTKEARSAGCIGNEFQCRPDGNCIPDLWRCDGEK 1077
QY 135 -----E---NGA---H-----T---I----- 141
Db 1078 DCEDGSEKCGNCTIRLCHDKTFRCSTGRCINNAWVGDDVDCBQDSEEDCDSEFLCG 1137
QY 142 -----AN-----NH----- 145
Db 1138 PPYPCANDTSVCLQPEKLCNCRKDCPDGSDGDLCECLNNGGCSNHCSSVVGRIVC 1197
QY 146 -----T---D-----MM---E-----VD-152
Db 1198 SCPEGHOLKDNRTCEIVDYCASHLRCQVCEQKHMVKCSYEGWALGTGDSCTSVDS 1257
QY 153 -----GD----- 154
Db 1258 FEAPIFSRHETRRIDLHGDDYLLVPLGRLNTIALDFHFNQSLLYWTDVVEDRIYRGKL 1317
QY 155 -----VE-----I-----I----- 157
Db 1318 SESGVSIAIEVVEHGLATEGLTVDMWAGNIYWIDSNLDQIEVSKLDGSLRATLAGAM 1377
QY 158 --P-----P-----N-----K----- 161
Db 1378 EHPRATALDPYGLFLEWTDWANDFPRIESASMSGACKTIYKDMKTGAWPNGLTVDHFER 1437
QY 162 -----AV-----VLRGHE-----SEV----- 172
Db 1438 RIWTDARSADIAYSAFVDGTNMIIEIRHGHEYLHSHPPFAVSLYSEVYVWTDWRTWTLAKAN 1497
QY 173 -----F-----I-----CA--176
Db 1498 WTCQNVSVIOKTSNAQPDLOIYHPSROPAAPNCAANEGRGPCSHLCLINHNESAACAP 1557
QY 177 -----W-----NP-----VS---D--- 182
Db 1558 HLMKLSDDKTCYEMKKFLLYARRSEIRGVDIDNPYVNFITAFVTPDIDDAVIDFDASE 1617
QY 183 --LL-----A---SG-----S-G---D-----S---T--- 192
Db 1618 ERLYWTDIKTQITRAFINGTGLETVISRDIQISIRGLAVDWVSRNLWYSSBDETFQINV 1677
QY 193 AR-----I-----W-----N----- 197
Db 1678 ARLDGSLKTSIIHGDKPQCLAAHPYRGKLYWTDGNTINMANMDGNSKILFQNKQEPVG 1737
QY 198 LS-----S---T-----SG-----S-----T----- 207
Db 1738 LSIYVYENKLYWISSGNGTINRCNLNDGNLEVIEMKEELTKATALTINDMKLWADQNL 1797

QY 208 -QL-----VLR-----H----- 213
Db 1798 AQLGTCKEDGRNPSILRNKTSQGVVHEMKYVDKEAQOGSNSQVNNGGCSQOLCLPTSETTR 1957
QY 214 -----C-----IR----- 216
Db 1858 TCMCTGVYVLQKNRMSCOGIESFLMYSVHEGIRGIPLEPRDKVDALMPISGAFAVGIDF 1917
QY 217 -----EG----- 218
Db 1918 HAENDTIYWDMLNKISRAKRDQTKWEDVVTNGLRVEGIAVDWIAGNIYWTDHGFNLI 1977
QY 219 -----GO-----D-----V 222
Db 1978 EVARLNGSFRYVLIISQLDQOPRSIAVHPEKGLFWTEWQVPCIGKARLDGSEKMWISV 2037
QY 223 ---P---S---NK-----D---V---T-----S-LD----- 232
Db 2038 GITWPNGISIDYBENKLYWCDAKSDKIERIDLTGANREVLSSGNSVDLFSVAVFGAYIY 2097
QY 233 W-----N-S-----E-CT----- 238
Db 2098 WSDRAHANGSVRRGHKNDAETVTMTGLGVNLKEIKIFNRVREKGTNVCAKENGCGQOL 2157
QY 239 -----LLA---T-----G-----S---Y--- 245
Db 2158 CLYGRNSRRTACAHGYLAGDGVTCRLRHEGYLLYSRTILKSIHLSDETNLNSPVRPYEN 2217
QY 246 -----D---G---F---A-----R---I----- 251
Db 2218 PNVFKNIIALAFDYNORREGTNRIFYSDAHFGNIQLIKONWEDROQVIVENVGSGEGLAYH 2277
QY 252 -----W-----TK---D---G-----NL----- 258
Db 2278 RAMDTLYWTSSTSTSTRHTVDTQTRFPAIDREAVITMSBDDHPHVALDECQNLFWTNW 2337
QY 259 -----A---ST-----LG-----Q-264
Db 2338 NEQHPISMRATLTGKNAHVVSVDILTPLNGLTIDHRAEKLYFSDGSLGKIERCEYDGSOR 2397
QY 265 H--K-GP-----IF-ALKW-----NK-----K----- 277
Db 2398 HVIKSGPGTFLSLAVDYSYIFWS-DWGRRAILRSNKYTGGETKILRSIPHOPMGIIAV 2456
QY 278 -----G-----N-----F 280
Db 2457 ANDTNSCELSPCALLGGCHDLCLLTPDGRVNSCRGDRVLLANNRCVTKNSSCNIIYSEF 2516
QY 281 -----I-----L-----S-AG-----V----- 286
Db 2517 ECGNGPCVDYVLTCDGI PHCKDKSDEKLLYCNENRSCSGFKPCYNRRRCVPHGKLCGTND 2576
QY 287 -----D-K---T-----T-I----- 291
Db 2577 CGDSSDELDCVKSTCSTVEFRCADGTICPRSARCNQNMDCSDASDEKGNNTDCTHFYKL 2636
QY 292 -----I-----W-----D-----AH----- 296
Db 2637 GVKSTGFIKRNSTSLCVLPSWICDGNDCGDYSDKLCQPVONKHKEENYFGCPSGRCIL 2696
QY 297 -T-----G-E-----A-----K----- 301
Db 2697 NTWVCDGQKCEBGLDELHCDSSCSWNQFACSVKVKCISKHWICDGEDDCGDSLDESISIC 2756
QY 302 -----Q-----Q-----F----- 304
Db 2757 GAVTCAADWFSCQGHACVPHQWLCDEGDCPDGSDDELSSAGCAPNNTCDENAFMCHNKV 2816
QY 305 --P--F--H-----S-----AP-----ALD-----VD--WQ-----S 318
Db 2817 CIPKQVCDHDDDCGSGSDEFTQCYGRQCGPEFRCA-DGRCLVNTLMQCDGDFDCPDSS 2875
QY 319 ---N-----NT--FASC-----STD-MC-I-----HV--C--- 334

Db	61	QSDSLDTCPEEVEIKPLNHIACHGSSACVHLSKLCNGVWDCPDGDFDEGHQCELLPSC	120
Qy	11	--L--V--Y--R--Y-L-Q--	17
Db	121	QQLNCQFKAMVRNATRCYCEDGEFAVEDGRSKDQDECSIYIGSCQCKNTYGSYACSC	180
Qy	18	-ESGF--S--H--SAF-T--F--GIK--S--H--	33
Db	181	VE-GYMQSDNRCKVKEPTDKAPMLLISSLETIELFYNGSKMTLSSANRNEIHTLD	239
Qy	34	--S--Q--I-S-QS--NI--	39
Db	240	FIYSEEMICWIESRESSNQLKCGQITKAGRLTDQRIINSLSQFQNVQMAFDWLTRNIYF	299
Qy	40	--NG--A-L--VP--	45
Db	300	VDHVSDFRIFVCFNFGSCVTLIESELHNPKAIAADPIAGKLPFTDYGNVPKIERCDLDGM	359
Qy	46	--P-A-AL--I--SIOKGLQ--Y-59	59
Db	360	NRTRIVYSKAEQPSALALDLNRLVWVLDYLVGVVYQGNRHTIYO-GRQVKHLYG	418
Qy	60	--V-E--AEVS--IN--DGR--E-68	68
Db	419	ITVFEDLYLA-TSSDNFNIRIRNFNGTDIHSIIKMSARGIYQKRTOPTVRSACEV	477
Qy	69	D--G--TL--F--DGR--76	76
Db	478	DAYGMPGCGSHICLLSSSYKTRCRTCRTGFNMGSDGRSKRPNKELFLFYGKGRPGIVRG	537
Qy	77	--PIE--SL--S-LI--84	84
Db	538	MDLNTKIADCEMPIENLNPRALDFAEANYFYADTTSFLIGROKIDGETERETILKDD	597
Qy	85	--D--AV--M--P--89	89
Db	598	LDNVEGIADVWIGNLWYNDGHRKTNVARLEKASQSRKTLLEGMSHPRAIVDPVNG	657
Qy	90	--D-V--V-Q--93	93
Db	658	WYWTWDWKEDKIDDSVGRIEKAWMDGVRQVFTSKMLWPNGLTDPHTSTLYWCDAYYD	717
Qy	94	--T-R--Q--96	96
Db	718	HIEKVFLNGTHRKVYSGKELNHPFGLSHHGNYVFWTDYMNYSIFQLDLMTNEVTLRHE	777
Qy	97	--QAY--R--D-K--LA--104	104
Db	778	RAPLFGLOIYDPRKQGDNWCRIINNGCGTLCIAIPAGRVACADNOLLDENGTCTFNP	837
Qy	105	--Q--Q--Q--107	107
Db	838	EEIRFHICKPGEFRCKNKHCIQARWKCDGDDCLDGSDSDSVTCFNHSCPDQFKQNNR	897
Qy	108	--A--AA--AA--AA--AA--116	116
Db	898	CIPKRWLCDGANDCGSNEDESNOTCTARTCOADQOFSCNGRCIPTAWLCDREDDCGDQTD	957
Qy	117	--AAS--Q--Q--121	121
Db	958	EVASCEPPTCEPLTQFICKSGRCSINXKHCDDTDDCGDRSDEVCGVHSCLDQDFRCSSGR	1017
Qy	122	--GS-A--KN-G--E-N-T--A--N--GE-134	134
Db	1018	CIPHWACDGNDCGDFSDETHINCTKEARSAPAGCIGNBFQCRPDGNCIPDLWRCDGEK	1077
Qy	135	--E--NGA--H--T--I--141	141
Db	1078	DCEDGSDEKGCNGTIRLCDHKTFPSCRSTGRGINNAWVGDDVDCEDQSDDEEDCDSPFCG	1137
Qy	142	--AN--NH--145	145
Db	1138	PPKYPCANDTSCVLOPEKLNCGRKDCPDGSDGDLCECLNNGCSNHCNVVPGRG1VC	1197
Qy	146	--T--D--MM--E--VD-152	152
Db	1198	SCPEGHQLKKNRTCBIVDYCASHURCSQVCSQKHVMKSCSYEGWALGTDGESCSTVDS	1257
Qy	153	--GD--154	154
Db	1258	FEAFIIFSIRHEIRRIDLHKGDYSLVPLGRNTIALDFHFNQSLLYWTDVVEDRIYRGKL	1317
Qy	155	--VE--I--157	157
Db	1318	SESGVSAIEVVVEHGLATPEGLTVDWIAAGNIYWIDNSLDQIEVSKLDSLRATLIAGAM	1377
Qy	158	--P--N--K--161	161
Db	1378	EHPRAIALDPYGIILPFTWDNDANFPRIEASMSGAGRKTIYKDMKTGAMPNGLTVDHFER	1437
Qy	162	--AV--VLRGHE--SEV--172	172
Db	1438	RIVWTDARSDAISAFYDGTNNMIEIIRGHEYLHPFAVSLYSGEVVYTDWRTWTLAKANK	1497
Qy	173	--F--I--CA--176	176
Db	1498	WTQONVSVIQTSAQPFDLQIYHPSPQAPNPCAANEGRGPCSHLCLINHNRSAAACACP	1557
Qy	177	--W--NP--VS--D--182	182
Db	1558	HLMLSSDKKCYEMKFLLYARRSEIRGVDINDPNVYFITAPTVPDIDDAVIDDASE	1617
Qy	183	--LL--A--SG--S-G--D--S--T--192	192
Db	1618	ERLYWTDIKTOITRAFINGTGLETVISRDQISIRGLAVDWYSRNLWYSSBEFDEQINV	1677
Qy	193	AR--I--W--N--197	197
Db	1678	ARLDGSLKTSIIHGIDKPOCLAAHPVRGKLYWTDGNTINMANMDGNSKILFQNOKEPVG	1737
Qy	198	LS--EN--S--T--SG--S--T--207	207
Db	1738	LSIDYVENKLYWISSNGTINRCLDGNLEVIEMKEELTKATALTIMDKKLWADQNL	1797
Qy	208	QL--VLR--H--213	213
Db	1798	AQLGTCNKRDRNPSTLRNKTSGVHMKYVDKEAQOQSGNSQVYNGGCSQCLPTSETTR	1857
Qy	214	--C--IR--216	216
Db	1858	TCMCTVGYTLQKNRMSCOGIESFLMYSVHEGIRGIPLEPRDKVDALMPISGAFAVGIDF	1917
Qy	217	--EG--218	218
Db	1918	HAENDTIYWDMLGINKISRAKRDQTWKEDVVTNGLRVRGIAVDIAGNIYTDHGFNLI	1977
Qy	219	--GO--D--V-222	222
Db	1978	EVARLNGSPRYVILISQGLDQPSIAVHPPEKFLFTWENGQVPCIGKARLDGSEKVMIVSV	2037
Qy	223	--P--S--NK--D--V--T--S-LD--232	232
Db	2038	GITWPNGISIDYBENKLYWCARSDKIERIDLDTGANREVLISGSNVDLFSVAVFGAYIY	2097
Qy	233	--N-S--E-GT--238	238
Db	2098	WSDRAHANGSVRRGHKNDAFETVTMTGLGNLKEIKIFNRVREKGTNVCAKENGCGQQL	2157
Qy	239	--LLA--T--G--S--Y--245	245
Db	2158	CLYGRNSRRTCAAHGYLAGDGVTCLEHGYLLYSGRITLKSITHLSDETNLNSVPVRYEN	2217
Qy	246	--D--G--P--A--R--I--251	251
Db	2218	PNYFKNIILAFDYNQRREGTNRIFYSDAHFGNIQLIKONWEDRQVIVENVGSVEGLAYH	2277

QY 252 W---TK---D---G---NL----- 258
D 2278 RAWDTLYTSSSTSSITRHTVDQTRFGAIDREAVITMSBDDHVLALDECONLMFTNW 2337
QY 259 A---ST---LG-----Q- 264
D 2338 NEQHPIMRATLTGKNAHVVSVDILTPNGLTIDHRAELKLYFSDGSLGKIERCEYDGSQR 2397
QY 265 H--K-GP---IF-ALKW---NK-----K----- 277
D 2398 HVIKSGPGTFLSLAVDSYIFWS-DWGRRAILRSNKYTGGETKIILRSIDIPHOPMGIIV 2456
QY 278 ---G---N---P 280
D 2457 ANDTNSCELSPCALLGGCHDLCLLTPDGRVNCSCRGDRVLLANNRCVTKNSSCNIYSEF 2516
QY 281 ---I---L---S-AG---V----- 286
D 2517 ECGNGDCVDYVLTCDGIPHCKDKSDEKLLYCNENRSCRGFKPCYNRRRCVPHGKLCDDTND 2576
QY 287 ---D-K---T---T-I---K----- 291
D 2577 CGDSSDELCKVSTCTVFERCADGTICIPRSARCNQNMDCSDASDEKGCNNTDCTHFKYL 2636
QY 292 I---W---D---AH----- 296
D 2637 GVKSTGFIKRNSTSLCVLPISWICDGSNDGDSDELKCPQNKHKCEENYFCGPGRCIL 2696
QY 297 -T---G-E---A---K----- 301
D 2697 NTWVCDGQKDCEDGLDELHCDSSCNQFACSVKCKISKHWICDGBDDCGDSDLESDSIC 2756
QY 302 ---Q---O---F----- 304
D 2757 GAVTCADMFSCGSHACVQPHWLCDGERDCPDGSDLSAGCAPNNTCDENAFMCHNKV 2816
QY 305 -P--F--H---S---AP---ALD---VD---WQ-----S 318
D 2817 CIPKQVCDHDDCGDGSDEFLOCGYRQCPEFRCA-DGRCLVTLWQCDGDFDCPDSS 2875
QY 319 ---N---NT-FASC---STD-MC-I-----HV--C--- 334
D 2876 DEAPINPRCSAEHSCNSFFM-CKNGRCIPS-DGLCDIRDGCGSDETNCHINECLSK 2933
QY 335 KL-G-QD-R-PI---KT---P---Q---G---H- 348
D 2934 KISGCSQDCQDLPVSYKCKWPGFQKDKGKTCVDIDECSSGFPSCQOCINTYGYKCHC 2993
QY 349 ---T---N----- 350
D 2994 ABGYETQDPNPGCRSLDEEPLIILAQHEIRKISTDGSNTYLLKQGLNNVIALDPDYR 3053
QY 351 -E-----V-NA-----I-K---W-D----- 358
D 3054 EBFYIWDSSRPNRGRINRMCLNGSDIKVYHNTAVPNALAVDWIGKNLYWSDTEKRIIEV 3113
QY 359 ---PT---GNLL-ASC----- 367
D 3114 SKLNGLYPTVLVSKRLKFRDLSDPRAGNLWIDCCPEYPIHGRVMDGTNOSVVIETKI 3173
QY 368 S-----D---M-----TL-K-I-W----- 376
D 3174 SRPMALTIDYVNHRLYWADENHIEFNSMDGSHRHKVPNQDIPQVIALTLFEDYIYWDGK 3233
QY 377 ---S-MK-----Q-D-----N---C----- 383
D 3234 TKLSRVHKTSGADRLSLNSWHAITDIQVHYSYRQPDVSKELCTVNNGGCSHLCLLPG 3293
QY 384 -VH-----D---L----- 387
D 3294 KTHTCACPTNFYLAADNRTCLSNCTASQPRCKTDKCIPIFWWKDVTDDCGDGSDEPDPC 3353

QY 388 ---Q---Q-----H-----NKE-I--- 394
D 3354 EFKCQPRFQCGTGLCALPAFICDGENDCGNSDELNCDFHVLCLAGQFKCTKNKKCIPVN 3413
QY 395 ---Y---T---I-K-W---SP----- 401
D 3414 LRCNGQDCGDEBEDEKOCPENSCSPDYFOCKTKKICISKLWVWDEDPDCADASDEANCDK 3473
QY 402 -T-GP---G---N---N-P----- 409
D 3474 KTCGPHFOCKNNCIPDHWRCNQNDQNSDSDCKPOTCTLKDFLCSNGDCVSRFW 3533
QY 410 ---N---A-----N----- 412
D 3534 CDGEFPCADGSDBKNCETSCSKDQFOCSNGQCLSAKWKCDGHEDCKYGEKNCPEAPPV 3593
QY 413 ---LMLA---SAS-----F----- 420
D 3594 CSSEYMCASGGCLSAKLCNKEPCVDCVSDGDEMDCVIECKEPOFOCKNKAYCIPIRWLCD 3653
QY 421 ---D-S---T---V-R---L-W-----D-V--- 429
D 3654 GIYDCVDSDEETCGGSGICRDBEFLNNLSLCKLHFW/CDGEDDCGDSDEAPDMCVKF 3713
QY 430 ---DR-----GI----- 433
D 3714 LCPPTPRYCRNDRICLQLEKICNGINDCGNSDEHSCGSLSKSPCKDBFTCSNRN 3773
QY 434 CI---HT---L-----T----- 439
D 3774 CIPMELQCDLDDCGDGSDEOGLKTPIEHTCENNNGPCDDAYCNOIKTSVFCRCCKPGF 3833
QY 440 ---K---H-----OE----- 443
D 3834 QRNMKGRECADNECLLFGICSHHCLNTRGSKVCVCDQNFQKNNSCIAKSGEDQALYA 3893
QY 444 ---P-----V-----Y----- 446
D 3894 NDTDILGFYVYPNYSGHQIISHVEHNSRI TGMVDVHYQRNVITWSTQFNPFGGIFYKMWIDA 3953
QY 447 ---S---VA-----FS----- 451
D 3954 REKQANSGLICPEFRPRDIAVDWVAGNVYTDHSHRMHFSYTYTHWTSLSRYSINVGOL 4013
QY 452 ---P-----DG----- 454
D 4014 NGPNCRTLTTNAGEPYAIVNPKGMVYTVIGDHSHTIEBAAMDGTLLRRLVQKNLQRP 4073
QY 455 ---R-Y---LA---S---GS----- 461
D 4074 TGLTVDFHGERIYWADFELSIIIGSVLYDGSFPVSVSSKQGLLHPHRIIDVFEDYIYGAP 4133
QY 462 ---F---DK----- 464
D 4134 KNGIFRVQFGHGSVEVLALGVDKTKSILVSHRYKQLNLPCLDLSCDFLCLNPSGAT 4193
QY 465 CV---H-----I-----W-----N--- 470
D 4194 CICPEGKYMNGTCHDDSLDSSCKLTCENGRCILNEKGLRCHCWPSYSGGRCEVNH 4253
QY 471 T---Q-----VC-L-----H-----Y----- 477
D 4254 SNYCQGGTCIPSTLGRPTCICALGFTGPNCKGKAVCEDSCHNGGSCVVTAGNPYCHCQA 4313
QY 478 -LNG---Q-V---L---L-N-----L-G-----R----- 488
D 4314 DYTGRQCYVYCHHYCVNSESCTIGNDGSVECVCPTRYEGPKCEIDKVCRRHGHCIINK 4373
QY 489 ---S---I---C-L---Y----- 493
D 4374 DNEDIFCNCTNKGIASSCOLCDGICYNGTGCOLDPETSIPVCVCTNWSGTQGERPAPKS 4433
QY 494 ---TL-----P 496

Db 4434 SKSEHISTRSIAIIVFLVLLVTLVTLVIGLVVCKRRKRTKTRRQPIINGINVEIGNP 4493
QY 497 -----H-H-----L-----V-----I-----PL-----504
Db 4494 SYNMYEVDHSDGGLLEPFSFMDPVKSRVYGGSSAFKLPHTAPPIYLSDLKGLPTFG 4553
QY 505 -----V-A-----L-I-----ELL-----V-LK 514
Db 4554 PTNYSNPVYAKLYMDQNCNRSNLASVDERKELLPKKIEIGIR 4595

RESULT 11

US-10-120-801-74
; Sequence 74, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spyttek, Kimberly
; APPLICANT: Mehraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glennnda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT FILING DATE: 2002-04-11
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 2001-04-24
; PRIOR FILING DATE: 2001-04-25
; PRIOR FILING DATE: 2001-04-25
; PRIOR FILING DATE: 2001-05-03
; PRIOR FILING DATE: 2001-05-16
; PRIOR FILING DATE: 2001-09-14
; PRIOR FILING DATE: 2001-09-14
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 74
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-120-801-74

Query Match 70.9%; Score 2588; DB 14; Length 5175;
Best Local Similarity 8.8%; Pred. No. 2.5e-29;
Matches 453; Conservative 49; Mismatches 8; Indels 4660; Gaps 389;

QY 1 M-----SI-----SS-----D-----E-----7
Db 1 MGRSPSWLYGVGLLLATTSSVNDKNDPTGKSSLAFAVFDITGSMFDDLQVREGAAK 60
QY 8 -----V-----N-----F-----L-----VY-----13
Db 61 IFKTVMAQREKLIYINYPHDFYLGELINTDSTYFMRLSKVYVHGGDCPEKTLTG 120
QY 14 -----R-----Y-----L-----QE-----SG-----F-----SHSAF-T 27
Db 121 ILKALQISLPSSFIYVFTDARSKDYHLEDEVLNTIQEKSSVVFVMTGDCGNRTHPGFRT 180

QY 28 -----FG-I-----KS-----H-ISO-----S-NI-----39
Db 191 YEKIAAASFGQVHLEKSDVSTVLEYVRHAVKQKVHLMYEAREGGTVSRNIPVDKHL 240
QY 40 -----N-----GA-----LV-----P-P-----46
Db 241 ELTISLSDKDDSDNLDIVLDRPEGRTVDKRLYSKEGGTIDLKKNVLIIRLKDPSFGVWTV 300
QY 47 -----A-----A-----L-I-----S-I-----52
Db 301 NTNSLKHTIRVFGHGAVDFKYGASRPLDRIELARPPVLNQDTYLLINNTGLIPPTV 360
QY 53 -----I-----Q-----KGL-----Q-Y-----VE-----AE-----63
Db 361 GEIDLVDYHSHLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYDNEVEFMRIAPT 420
QY 64 -----V-----S-I-----N-----ED-----GTLF-----73
Db 421 AIGSVIVGGPRAFMSPHQEFVGRDLNLCTVESASAVTIYWKGTGEDIIGPLFYHNTD 480
QY 74 -----D-----G-----R-----P-I-----E-----79
Db 481 TSVWTIPELSLKDAGEYECRVISNNGNYSVKTRVETRESPPEIFGVRNVSVPLGEAAFLH 540
QY 80 -----S-----L-----SL-I-----DA-----86
Db 541 CSTRSAGEVEIRWTRYGATVFNPNPTNPTNGTLKIHVTRADAGVYECMARNAGCMST 600
QY 87 -----VM-----P-DV-----V-----92
Db 601 RKMRLDIMEPPSVKVPQDVYFNMREGVNLSCAEMGDPKPEVHVYFKGRHLLNDYKQVG 660
QY 93 -----T-----R-----Q-----QAYRD-----KLA-----Q-----105
Db 661 QDSKFLYIRDTATHDEGTGECRAMSQAOA-RDITDMLATPPKVEIIONKMMVGRGDRV 719
QY 106 -----Q-----Q-----A-----A-----A-----A-----111
Db 720 SPECKTIRGKPHKIRWFKNGKDLKPDYIKINEGQLHMGAKDEBAGAYSCVGNMAG 779
QY 112 -----A-----A-----A-----A-----A-----A-----114
Db 780 KDQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEMQKGNVLLATLNPP 839
QY 115 -----A-----A-----A-----ASQ-----120
Db 840 RYTQADGNLLITDAQIEDQGOFTCIARTYQQSQSSTLMVTLGLVSPVLGHVPPEEQLI 899
QY 121 -----Q-----GS-----123
Db 900 EQDLTLSCVVVLGTGPKPSIVMKDDKPVEEGPTIKIEGGGSLRLRGNGPKDEGKYTCI 959
QY 124 -----A-----K-----N-----N-----GE-----128
Db 960 AVSPAGNSTLHINVLKPKPEFVYKPEGGIVFKPTISGMDEKHAHVAVNSTHVDLDEGFA 1019
QY 129 -----NT-----A-N-----G-----133
Db 1020 IPCVSGTTPPIITWLDGRPITPNSRDFVTADNTLIVRKADKSYSGVYTQATNSAGD 1079
QY 134 -----E-----E-----N-----136
Db 1080 NEQKTTIRIMNTPMISPGQSSFMNVVDDLFTIPCDVYGDPKPVITWLLDDKPFTEGVNNE 1139
QY 137 -----G-----AH-----TI-A-N-----N-----144
Db 1140 DGSLLTIPNVNEARHGTFTCHAQNAAGNDRFTVTLTVHTTPTINAENQEKIALQNDIVLE 1199
QY 145 -----HT-----D-----MME-----V-----D-----152
Db 1200 CPAKALPPPVRLWYSEKIDSLIPIHTIREDGALVQLNVKLENTGVFCQVSNLAGEDS 1259
QY 153 -----G-DV-----EIP-----P-----NK-AV-----163

Db 1260 LSYTLTVHEKPKIIESEVGVVDVVKGFTTIBIPCRATGVPVIRTNKNGIDLWDEKFS 1319
QY 164 V-----LR-----G-----H-----E-----S----- 170
Db 1320 VDNLTGTLRIYEADKNDIGNVNCVVTNEAGTSQMTTHVDVQEPPIILPSTQNTNNAVUGDR 1379
QY 171 -----E-----V-----F-----I-----C----- 175
Db 1380 VELKCYVEASPASVTFRRGIAIGTDTKGYYVESDGLTVIQSASVEDATIYTCASNPA 1439
QY 176 -----A----- 176
Db 1440 GKAEANLQVTIASPIKDPDVTQBSIKESHPPSLYCPVFSNPLPQISWYLNKPLIDD 1499
QY 177 -----W----- 178
Db 1500 KTSWKTSDDKRLHVPKAKITDSGVVKCVARNAGSGSKSFQVEVIVPLNLDSEYKKKV 1559
QY 179 -----PVS----- 181
Db 1560 FAKEGEEVTLGCPVSGFPVQPINNVVDGTWVEPGKYGATLSNDGLTLHFDVSYSVKQEG 1619
QY 182 -----D-----L-----LA-----S 186
Db 1620 NYHCVAQSGNILDIDVELSVLAVPIVGEDDNLEVLGKDISLSCDLQTESDDKTTFFWS 1679
QY 187 -----GS-----G-----DS-----T-----A-----R----- 194
Db 1680 INGSSEDRPDNVQIPSDGHRLYTDAKENNGKCMRVNNSAGKAERTLTLOVLEPPV 1739
QY 195 -----EN----- 199
Db 1740 EPVFEANQKLGNNPIILQCVTGNPKPTVIWKIDGNDVDKSWLPDESLSLLRIEKLTKG 1799
QY 200 -----EV----- 207
Db 1800 SAQISCTAENKAGTASRDFFIQIAAFTFKNEGDOETIFRESEITLDCPVSIGDPQITW 1859
QY 208 -----Q-----L-----VL-----R-----H-----C----- 216
Db 1860 MKQGLPLTENDAIFLTDNLRLTILNANRDHEDIYTCVANTAGQVSKDFVGVVQLPKIK 1919
QY 217 -----EG-----G-----Q-----DVP----- 223
Db 1920 NAVVTLIINEGERIILTCDAGNPTTAKWDFNQGLDPKEAVFVNNHTVWVNNVTKYHT 1979
QY 224 -----SNK-----DV-----T-----S-----L-----D----- 232
Db 1980 GVKYCATNKVQAVKTINVHRTKPRFESGLTESELTVNLTRSITLECDVDDAIGVGIS 2039
QY 233 W-N-----SE-----G-----TL-----L-----A-----T-----GSY----- 245
Db 2040 WTVNGKPLAETDGVQTLAGRFLHIVSAKTDHDSYACTVTNEAGVATKTNLFVQVPP 2099
QY 246 -----DG----- 248
Db 2100 TIVNEGGEYTVIENNSVLVPCVETGKPNVPTWKDGRPVGDLKSVQVLSGQQKIVHA 2159
QY 249 -----A-----R----- 250
Db 2160 EIAHKGSIYICMAKNDVGTAISPDVDIITRPMIQIKNIIVTAIKGALPFKCPIDDDKN 2219
QY 251 -----IW----- 257
Db 2220 FRQGIWLNRNYQPIDLEADARITLSNDRRLTILNVNTEDEGQVSCRVRNDAGENSFDF 2279
QY 258 -----L-----A-----S-----TL----- 265
Db 2280 KATLVPPPIIMLDKDKNTAVEHSTVLTSCPATGKPEPDITWFKDGEAIHIENTADIIP 2339
QY 266 ----- 265

Db 2340 NGBELNGNQLKITRIKEGDAGKYTCESADNSAGSVQDVNVNVNVTIPKIEKDIPSDYESQQ 2399
QY 266 -----K-----G-----P----- 269
Db 2400 NERVVISCVPYARPPAKITWLKAGKPLQSDKFVKTSANGOKLYLFKLRETDSSKYTCIAT 2459
QY 270 -----F-----A-----L-----K-----W----- 274
Db 2460 NEAGTDKDPKVSMLVAPSFDEPNVRRITVNSGNPSTLHCPAKGSPSPITITWLKGNAI 2519
QY 275 -----N-----K----- 276
Db 2520 EPNDRYVFFDAGRQLQISKTEGSDQGRYTCIATNSVSGDLENTLVBVIPPVIGERREA 2579
QY 277 -----K-----G-----N-----FI-----LS----- 283
Db 2580 VAVIEGSELSFCDNSNSTGVDVWQXGLTINQDTLGRGDSFQIPSSGKMSFSAKSD 2639
QY 284 -----AG-----V-----DK-----T-----TI-----I----- 292
Db 2640 SGRYTCIVRNPAEARKLFDFAVNDPPSISDELSSANIQTIVPYYPVEINCVVSGSPHPK 2699
QY 293 -----W-----D-----A-----H-----T-----G----- 298
Db 2700 VYWLFDKPLEPDSAAVELTNNGETLKIYRSQVEHAGTYTCEAQNNGVARKDFLVRVTA 2759
QY 299 -----E-----AK-----Q----- 302
Db 2760 PPHEKEREVARVGDVMTLLTCNAESSVPLSSVYVWHAHDESQVNGVITSKYAANEKTLN 2819
QY 303 -----Q-----F----- 304
Db 2820 VTNIQLDDBGFYCTAVNEAGITKFKFLIVETPYFLDQOKLYPIILGKRLTDCSATG 2879
QY 305 -----P-----F----- 306
Db 2880 TTPPTILFMKQKRLNESDEVDIIGSTLVIDNPQKEVEGYTCIAENKAGRSEKMMVEV 2939
QY 307 -----H----- 308
Db 2940 LLPPLKSKWENVEVQAGDPLTLECPEDTSGVHTWSRQFGKQDGMRAQSSDKSKL 2999
QY 309 -----A-----P-----A-----L-----D-----V-----D----- 315
Db 3000 YIMQATPEDADYSCTIANDAGAEAVFQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059
QY 316 -----W----- 317
Db 3060 EGIPPEISWFLDGKPILEMPGVYTKQDLSLRIDNIKNQEGRYTCVAENKAGRAEQDT 3119
QY 318 -----S-----N-----N-----N-----T-----F----- 322
Db 3120 YVEISEPPRVNASEVMRVVEGRQTTIRCEVFGNPEPVVNWLDKGPYTSDDLQSFSTKLS 3179
QY 323 -----A-----SC-----S-----TD-----N-----C----- 330
Db 3180 YLHLRETTLDGGTYTCIATNKAGESQTTDDEVLPVPRIEDEERVLQKEGTYNVVHCQ 3239
QY 331 -----I-----HV-----C-----K 335
Db 3240 VTGRVPVYVTKRNGKEIEQFPNVLHNRATRADEGKYSIASNAGTAVADFLIDVFTK 3299
QY 336 -----L-----Q-----D-----RP-----I-----K----- 343
Db 3300 PTFETHETFTNIVEGESAKIECKIDGHPKPTISLWKGGRPFNNDNIIILSPRGDTMLKA 3359
QY 344 -----T-----F-----Q----- 346
Db 3360 QRFDGLYTCVATNSVGDSEQDFKVNVTYKPYIDETIDOTPKAVAGGEIILKCPVLGNPT 3419
QY 347 -----G-----HT-----N-----E-----VN-----A-----I----- 355
Db 3420 PTVTWKRGDDAVPNDNRHTIVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVIG 3479

QY 356 ---K---W---D---P---T---360
Db 3480 KPTFRKGNLYEVIENDTTIMDCGVTSRPLSPISWFRGDKPVLYLDYSISPDGSHITI 3539
QY 361 ---G---N---L---LA---S---C---S 368
Db 3540 NKAKLSDGGYICRASNEAGTSDIDLILKLVPPKDKSNIGNPLAIVARTIYLECPIS 3599
QY 369 ---D---D---M---371
Db 3600 GIQPDVIMTKNGMINDMDSRVILAQNNETFGIENVQVTDQGRYTCATNRGKASHDF 3659
QY 372 ---TL---K---I---W---S---M---378
Db 3660 SLDVSPPEFDIHGTOPTIKREGDTITLCPKLAEDIAQVMDVSVTKDSRALDGLTD 3719
QY 379 ---K---Q---D---381
Db 3720 NVDISDGRKLTISOASLENAGLYCTALNRAGEASLEFKVEILSPPVIDISRNDVQPV 3779
QY 382 ---N---C---V---H---D---L---Q---Q---H---380
Db 3780 AVNQPTIMRCVATGHPSPS IKMLKNGKEVTDDEINIRIVEQGVQLILRTDSDHAGKWSVCV 3839
QY 391 ---N---KE---I---Y---TI---KW---399
Db 3840 AENDAGVKELEMLVDVFTPPVSVKSDNP IKALGETITLFCNAGNPPYQPKWAKGGLI 3899
QY 400 ---SP---T---G---P---GT---N---407
Db 3900 FDSFDGARISLKGARLDIPLHKKTVDGDTYCOALNAAGTSEASVVDVLPVPPINRDGID 3959
QY 408 ---N---P---409
Db 3960 MSPRLPAQSLTLQCLAQGPVQMRWTLNGTALTHTSTPGITVASDSTFTIQINNVSLSDK 4019
QY 410 ---N---A---NLM---LA---S---417
Db 4020 GVYTCYABNVAGSDNLMYVVDVQAPVISINGTKQVIEGELAVIECLVEGYPAPQVSWLR 4079
QY 418 ---A---SPD---421
Db 4080 NGRVETGVQVRYVTDGRMLTIEARSLDSGIYLCSATNEAGSAQOAYTLEVLVSPKII 4139
QY 422 ---ST---VR---L---W---DV---429
Db 4140 TSTPGVLTPSSGSKFSLPCAARGYPDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKABE 4199
QY 430 ---D---R---431
Db 4200 RHLIYEAKNDAGADTLEFPVQTVIAPKISTSGNRYINGSEGTETVIKIESESSEFS 4259
QY 432 ---G---I---C---434
Db 4260 WSKNGVPLLSNNLIFSEDIYKILSTRLSQGEYSCTAANKAGNATQKTNLVGAPK 4319
QY 435 I---H---TL---TK---440
Db 4320 IMERPRTQVHVKGQDQVTLWCEASGVQPAITWYKDNELLTNTGTGDETATTKKKSVPFSSI 4379
QY 441 ---440
Db 4380 SPSQAGVYTCKAENWVASTBEDIDLIVMIPPEVVPVPMNVSTNPTQTVFLSCNATGIPEP 4439
QY 441 ---H---Q---442
Db 4440 VISWRDSNTAIQNEKYQILGTTLATRNVLPPDDGFYHICIAKSDAGCKIATKRLVKNKP 4499
QY 443 ---E---P---V---Y---446
Db 4500 SDRPAPIWVECKEKGPKKTEYIMIDRGDTDPDNPQLLPWKDVEDSSLSNGSIAYRCMPGPR 4559

QY 447 -S--V--A--F--S--P--DG--454
Db 4560 SRTVLLHAAPQFIYVVKPKTTAAIGAIVELCSAAGPHTTITWAKDGKLIEDSKFEIAY 4619
QY 455 ---454
Db 4620 SHLKVTLNSTSDSGEYTCMAQNSVSGSVTSVAFINVDNNILPTPKPSSNQKNVAITCYER 4679
QY 455 ---R---Y---LA---S---G---460
Db 4680 NOAYSRLTWXNGVPMKPLAGIHFPMNNGSLVILDTSSLKEGDELEYTKVNRHRSI 4739
QY 461 ---S--P---DK---C--V---H--IW---469
Db 4740 PHLTSAFEGVPEVKTIDKVEVNGDSVILDCEVTSPLTHVVVTKNDQKMLDDDAIYVL 4799
QY 470 ---469
Db 4800 PNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQLNVYGGSSRRREAYKKENEDASTTIT 4859
QY 470 ---N---TO---472
Db 4860 TTSPTTTTTEPLTTTIIIPALITLPAKOYPTDDYHEGSANDDGGFTTQDSLFEFNPPLH 4919
QY 473 ---V---C---L---H---476
Db 4920 PEISVNTTCAGTINENGDCVCKDKTHNLKILTGENHCPEGFAMPNPHTRICEDLDECAF 4979
QY 477 Y---LN---G-QVLLN---LG---R---SI-C---L---492
Db 4980 YQPCDFECINYDGGFQ--CNCPLGYELABEGCRDVNCESVRCEDGKACFNQLGGYECID 5037
QY 493 ---YTL---P--H---H---L---V---500
Db 5038 DPCFANYSLVDDRYCEPCENCTSTPIQVHMLAIPSLPISHIATILATYDKSGRVLNDTT 5097
QY 501 -VI---PL---V--A---L---I---EL--510
Db 5098 YALSDGAFGLARGWTSFGPTIKAVRGHAQVWTVNRVLAAGDHHKVRVRAHSDHATNELH 5157
QY 511 ---LVL 513
Db 5158 APKETNPLVL 5167

RESULT 12

US-10-369-493-6859
; Sequence 6859, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6859
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6859

Query Match 70.9%; Score 2588; DB 14; Length 5175;
Best Local Similarity 8.8%; Pred. No. 2.5e-29;
Matches 453; Conservative 49; Mismatches 8; Indels 4660; Gaps 389;

QY 1 M-----SI-----SS-----D-----E----- 7
Db 1 MGRSPSWLYGLVGLLLATTSSVNDKNDPTGKSLAFVFDITGSMFDDLVOVREGAAK 60
QY 8 ---V-----N-----F-----L-----V----- 13
Db 61 IPKTVMAQREKLIYNYIMVPHDPYLGEIINTDSTYFMQLSKVYVHGGDCPEKTLTG 120
QY 14 ---R-----Y-----L-----QE-----SG-----F-----SHSAF-T 27
Db 121 ILKALQISLSPSSFYVFTDARSXDYHLEDEVLNTIOEKQSSVVFVMTGDCGNRTHPGFRT 180
QY 28 ---FG-I-----KS-----H-----ISQ-----S-----NI----- 39
Db 181 YEKIAAASFGQVHFLEKSDVSTVLEYVRHVAVKQKHLMYEARERGCTVSRNIPVDKHL 240
QY 40 ---N-----N-----GA-----LV-----P-----P----- 46
Db 241 ELTISLGDKDDSNLDIVLRDPGRVTKRLYSKEGGTIDLKNVKLIRLKDPSGWTV 300
QY 47 ---A-----A-----A-----LI-----S-----I----- 52
Db 301 NTNSRLKHTIRVFGHGADEFKYGFASRPDLRIELARPPVNLQDVTYLLINWTGLIIPGTV 360
QY 53 --I-----Q-----KGL-----Q-----Y-----VE-----AE----- 63
Db 361 GEIDLVDYHGSLYKAVASPHRTNPNMYFAGVPFPPKGLFFVRVQGVDEDNVEFMRIAPT 420
QY 64 ---V-----S-----I-----N-----ED-----GTLF----- 73
Db 421 AIGSVIVGPPAPMSPIHQBFGVGRDLNLCTVSASAYTIYVWKTGEDITIGSLFYHNTD 480
QY 74 ---D-----D-----G-----R-----P-----I-----E----- 79
Db 481 TSVWTIPELSLKDAGEYECRVISNNGYNSVKTRETRESPEIFGVRNVSVPLGEAAFLH 540
QY 80 ---S-----L-----SL-----I-----DA----- 86
Db 541 CSTRSAGEVEIRWTRYGATVFNPNTPNTPNTGTLKIHVTRADAGVYECMARNAGMST 600
QY 87 ---VM-----P-----DV-----V----- 92
Db 601 RKMRLDIMEPPSVKVTPODYFNMREGVNLSCBAMGDPKPEVHWYFGRHLLNDYKYQVG 660
QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q----- 105
Db 661 QDSKFLYIRDATHDEGTIECRAMSAQQA-RUTDMLATPPKVEIIQNMVMGRGDRV 719
QY 106 ---Q-----Q-----Q-----A-----A-----A-----A----- 111
Db 720 SPECKTIRGKPHKIRWFKNKGDLIKPDYKINEGQLHMGAKDEADAGAYSCVGENMAG 779
QY 112 ---A-----A-----A-----A-----A-----A----- 114
Db 780 KDQVANLSVGRVPTTIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNP 839
QY 115 ---A-----A-----A-----ASQ----- 120
Db 840 RYTQLADGNLLITDAQIEDQGQFTCIARNYTGQSSQSTLMTVGLVSPVLGHVPEEQLI 899
QY 121 --Q-----Q-----GS----- 123
Db 900 EGQDLTLCVWVLTGPKPSIVMIKDKPVBEETIKIEGGSSLLRLRGNPKDEGKYTCI 959
QY 124 ---A-----A-----K-----N-----GE----- 128
Db 960 AVSPAGNSTLHINVLKPKPEFVYKPEGGIVFKPTISGMDEKHAVVVNSTHVDLDEGEPA 1019
QY 129 ---NT----- 248
Db 1020 IPCWSGTPTPIITWYLDGRPTPNRSRDTFTVADNTLIVRKADKSYSVYTCQATNSAGD 1079
QY 134 -E-----E-----N----- 136

Db 1080 NEQKTTIRIMTWMIPGQSSFNWVDDLTFTPCDVGYPKPVITWLLDDKPFTEGVVNE 1139
QY 137 -G-----AH-----TI-----A-----N----- 144
Db 1140 DGSLLTIPNVNEARHGFTTCHAOAAGNDTRVTTLTHTPTTINAEOEKIALQNDIDVLE 1199
QY 145 ---HT-----D-----MME-----V-----D----- 152
Db 1200 CPAKALPPPVLWYBEGEKIDSQIPIHTTREDGALVLQNVKLENTGVFVQCVSNLAGEDS 1259
QY 153 ---G-----DV-----EIP-----P-----NK-----AV----- 163
Db 1260 LSYTLVHEKPKIIEVPGVVDVVGKFTTIEIPCRATGVPVIRTNKNGIDLKNDEKKFS 1319
QY 164 V-----LR-----G-----H-----E-----S----- 170
Db 1320 VDNLGLTIRIYEADKNDIGNYCVVNTNEAGTSQMTTHVQOEPIILPSTQNTNTAVVGDR 1379
QY 171 ---E-----V-----F-----I-----C----- 175
Db 1380 VELKCYVEASPPASVTVFRGIAIGTDTKGYYVESDGLTVIQSASVEDATIYTKASNPA 1439
QY 176 ---A----- 176
Db 1440 GRAEAMLQVTVIASPDIKDPDVVTQESIKESHPPFSLYCVPFSNPLPQISWYLANDKPLIDD 1499
QY 177 --W----- 178
Db 1500 KTSWKTSDDKRLHVFKAKITDSGVYKVCARNAAGBSKSFQVEVIVPLNLDSEKYYKKV 1559
QY 179 ---PVS----- 181
Db 1560 FAKGSEVTLCGPVSGFPVPOINWVVDGTVVEBKKYKGTATLNDGLTLHFDVSVKQEG 1619
QY 182 ---D-----L-----LA-----S----- 186
Db 1620 NYHCVASKGNILIDIVELSVLAVPIVGDNDLNVFLGKDISLSCDLQTESDDKTFVWS 1679
QY 187 --GS-----G-----DS-----T-----A-----R----- 194
Db 1680 INGSSEDRPDNVQIFSDGHRLYITDAKPNNGKMYCRVTNSAGABRTTLTDLVEPPVFV 1739
QY 195 ---IW-----N-----LS----- 199
Db 1740 BPVFEANQKIGNNPIILOQCQVGNPKPTVIWKIDGNDVDKSWLFDSESLRLIEKLTGK 1799
QY 200 ---EN-----S-----T-----S-----GS-----T----- 207
Db 1800 SAQISCTAENKAGTASRDPFIQNIAPTFKNEGDQETIFRESEITITLDCPVSIGDQITW 1859
QY 208 --Q-----L-----VL-----R-----H-----C-----IR----- 216
Db 1860 MKQGLPLTENDAIFTLDNTRLTILNANRDHDIYTCVANNTAGQVSKDPDVVVQVLPKIK 1919
QY 217 ---EG-----G-----Q-----DVP----- 223
Db 1920 NAVTLEINEGEEIILTCDAEGNPPTAKWDFNQGLPKAEVFNNNHTVVNNVTKYHT 1979
QY 224 ---SNK-----DV-----T-----S-----L-----D----- 232
Db 1980 GYVKCATNKVQAVKTIINVHVTKPRPESGLTESELTNLTFRSITLEDVDDAIGVGIS 2039
QY 233 W-----N-----SE-----G-----TL-----L-----A-----T-----GSY----- 245
Db 2040 WTVNGKPLAETDGVQTLAGGRFLHVSAKTDDHGSYACTVTNEAGVATKTNLFVQVPP 2099
QY 246 ---DG----- 248
Db 2100 TIWNEGVEYTIENNSLVLPCBVTGKPNPVVTTWKDGRPVGLKSVQVLSGQQPKIVHA 2159
QY 249 ---A-----A-----R----- 250

Db 2160 EIAHGSYICMAKNDVGTAEISFDVDIITPEMIQKGIKNIVTAIKGALPFKPCIDDDKN 2219
Qy 251 ---IW---T---K-D-G-N---257
Db 2220 FKQIILWRYQPIDLEADARITRLSNDRRLTILNVNTEDEQYSCRVKNDAGENSFDF 2279
Qy 258 ---L---A-S-TL---GO-H---265
Db 2280 KATVLVPPTIIMLDKOKNKTAVEHSTVTLSCPATGKPEPDITWFKDGEALHNIADIIP 2339
Qy 266 ---265
Db 2340 NGELNGNQLKIITRIKEDGACKYCEADNSAGSVEQDVNVNVIITPKIEKGIPSDYESQ 2399
Qy 266 ---K-G-P---I--269
Db 2400 NERVWISCPYARPPAKITWLKAGKPLQSDKFVKTSAHQGLYLFKLRETDSSKYTCIAT 2459
Qy 270 ---F---A---L---K---W---274
Db 2460 NEAGTKRDPKVMVLVAPSPDEPNIVRRITVNSGNFSTLHCPAKGSPPTIITWLDGNAL 2519
Qy 275 ---N---K---276
Db 2520 EPNDRYVFFDAGRQLQISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPVVIDGERREA 2579
Qy 277 ---K-G-N---FI---LS---283
Db 2580 VAVIEGFSSELCDSNSTGVDEWQKDLTINQDTLRGDSFIQIPSSGKMSFLSARKSD 2639
Qy 284 ---AG---V---DK-T---TI---I---282
Db 2640 SGRYTCIVRNPAEARKLFDFAVNDPPSISDELSSANIQTIVYYPVEINCVVSGSPHPK 2699
Qy 293 ---W---D-A---H---T---G---298
Db 2700 VYMLFDDKPLEPDSAAVELTNGGETLKI VRSQVEHAGTYTCEAQNNGVKARKDPLVRVTA 2759
Qy 299 ---E---AK---Q---302
Db 2760 PPHEKEREVARVGTMLLTCAESSVPLSVYVWHAHDESQNGVITSKYAANEKTLN 2819
Qy 303 ---Q---F---F---304
Db 2820 VTNIQLDDEGYCYCTAVNEAGITKKFKLIVETPTPYLDQOKLYPIILGKRLTLDCSATG 2879
Qy 305 ---P---F---306
Db 2880 TPTPTILFMKDGKRLNESDEVDIITGSTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVEV 2939
Qy 307 ---H---S---308
Db 2940 LLPKLSKEWINVEVQAGDPLTLECPIDTSGVHITWSRQFGKDGQLDMRAQSSDKSL 2999
Qy 309 ---A-P-A-L---D---V-D---315
Db 3000 YIMQATEDADSYCIAVNDAGAEAVQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059
Qy 316 ---W---Q---317
Db 3060 EGIPPEISWFLDKPILEMPGVTYKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQDT 3119
Qy 318 ---S---N---N---T---F---322
Db 3120 YVEISEPPRVMASEVMRVEGRTTIRCEVFGNPEPVMWLKDGEPYTSLLQFSTKLS 3179
Qy 323 ---A---SC---S---TD---M-C---330
Db 3180 YLHLRETTLDAGGYTCIATNKAGESQTTTIDVELVPPRIEDEBRLVQKEGNTYMHQC 3239
Qy 331 ---I---HV---C---K---335
Db 3240 VTGRPVVYTWKRNKEIQENFVLHIRNATRADEGKYSCIASNEAGTAVADFLIDVFTK 3299

Qy 336 ---L-CQ---D---RP---I-K---343
Db 3300 PTFETHETFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMNDNIIILSPRGDTLMLKA 3359
Qy 344 ---T---F---Q---346
Db 3360 QRFDGGLYTCVATNSVGDSEQDFKVVVTKPYIDETIDQTPKAVAGEIILKCPVLGNPT 3419
Qy 347 ---G---HT---N---E---VN---A-I---355
Db 3420 PTVTWKRGDDAVPDSRHTIVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVIG 3479
Qy 356 ---K---W---D---P---T---360
Db 3480 KPTFVRKGNLYEVIENDTIITMDCGVTSRPLPSISWFRGDKPVLVYDRYSISPDGSHITI 3539
Qy 361 ---G---N---L---LA---S---C-S---368
Db 3540 NKAKLSGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIICNPLAIVARTIYLECPIS 3599
Qy 369 ---D---D-M---371
Db 3600 GIQPDVITWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCIATNKGKASHDF 3659
Qy 372 ---TL---K---I---W---S---M---378
Db 3660 SLDLVSPPEFDHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLTD 3719
Qy 379 ---K---Q---D---381
Db 3720 NVDISDGRKLATISQASLENAGLYTCIALNRAGEASLEPKVEILSPPPVIDISGRNDVQOV 3779
Qy 382 ---N---C-V-H---D---L---O-Q---H---390
Db 3780 AVNQPTIMECAVTHGPPPSIKWLKNGKEVTDENIRIVEQGQVQLRLTDSHAGKWSVCV 3839
Qy 391 ---N---KE---I-Y-TI---K---399
Db 3840 AENDAGVKELEVLVDVFTPPVSVKSDNPIKALGETITLFCNAGSNPYQLKWKAGGSLI 3899
Qy 400 ---SP---T-G---P---GT---N---407
Db 3900 FDSPDGARISLKGARLDIPLHKKTDVGYTCOALNAAGTSEASVSVDVLVPPPEINRDGID 3959
Qy 408 ---N---P---409
Db 3960 MSPRLPAQOSLTLQCLAQKPVQMRWTLNGTALTHTSTGITVASDSTFIQINNVSLSDK 4019
Qy 410 ---N-A---NLM---LA---S---417
Db 4020 GVTYCAENVAGSDNLMYVNVVQAPVISNGGTQKQVIEGELAVIECLVEGPAQVSWLR 4079
Qy 418 ---A-SFD---421
Db 4080 NGNRVETGVQVRYVTDGRMLTIIIEARSLDSGIYLCSATNEAGSAQOAYTLEVLVSPKII 4139
Qy 422 ---ST---VR-L---W---DV---429
Db 4140 TSTPGVLTSSSGKFSPLCAVRCYPPPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4199
Qy 430 ---D---R---431
Db 4200 RHLIYECTAKNDAGADTLEFPVQTI VAPKISTSGNRYINGSEGTETVIRKEIESESEFS 4259
Qy 432 ---G---I---C---434
Db 4260 WSKNGVPLLPNNLIFSEDYKLIKILSTRLSDOGEYSCTAANKAGNATOKTNLVGVAPK 4319
Qy 435 I---H---TL---TK---440
Db 4320 IMERPRTQVVKGDQVTLCEASGVQPALITWYKDNELLTNTGVDETTATTKKSVIFSSI 4379

441 QY ----- 440
4380 Db SPSQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMNVSTNPNROTFLSCNATGIPEP 4439
441 QY ----- 442
4440 Db VISWRDSNTAIONNEKYQILGTTTLAIRNVLPDDGGFYHCIAKSDAGQKIATKLIWKP 4499
443 QY ----- 446
4500 Db SDRPAPIWECDEKGPKEKTEYIMIDRGDTPDDNPQLLPWKDVEDSSLSNGSIARCMGPR 4559
447 QY -S-V-A-F-----P-----DG----- 454
4560 Db SSRTVLLHAPOFIVKPKNTTAAIGAIVELRCSAAGPPHTITWAXDGKLIEDSKPEIAY 4619
455 QY ----- 454
4620 Db SHLKVTLNSTSDGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAITCYER 4679
455 QY -R-Y-LA-----S-G----- 460
4680 Db NOAYSRLTWEYNGVMPKNLAGIHFMNNGSLVILDTSSLSKGLDLELYTCKVNRRRHSI 4739
461 QY -S-F-----DK-----C-V-----H-W----- 469
4740 Db PHLTSAFEGVPEVKTIDKVEVNGDSVLDCEVTSPLTTHVVTWKNDQKMLDDDAIYVL 4799
470 QY ----- 469
4800 Db PNNSLVLLNVEKYKCVASNSIGKAFDDTQLNVYGGSSRRAYKKENEDASTTTIT 4859
470 QY ----- 472
4860 Db TTSPTTTTETPLTTIIPALITLPAKQYPTDDYHEGSANDDGGFTTQSLPEFNPPLH 4919
473 QY -V-----C-----L-----H----- 476
4920 Db PEISVNTDCAGTINENGCDVKDKGTHNLKILGTENHCPEGFAMNPHTRICEDLDECAF 4979
477 QY -LN-G-QVLLN-LG-----R-SI-C-----L----- 492
4980 Db YQPCDFECINYDGGFQ-CNCPGLGYELABEGCRDVNECESVRCEGKACFNQLGGEYCID 5037
493 QY -YTL-----P-H-----H-L-----V----- 500
5038 Db DPCPANYSLVDDRYCEPECENTSTSTPIQVHMLAIPSGLPISHIATLTAVDKSGRVLNDTT 5097
501 QY -VI-----PL-----V-A-----L-----I-----EL- 510
5098 Db YALSDTGAPLARGMTSGPPTIKAVKRGHAQVWTVNRVLAAGDHHKVRVRAHSDHATNELH 5157
511 QY -----LVL 513
5158 Db APKRETNFLVL 5167

RESULT 13

US-10-369-493-6861
; Sequence 6861, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6861
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6861

Query Match 70.9%; Score 2588; DB 14; Length 5175;
Best Local Similarity 8.8%; Pred. No. 2.5e-29;
Matches 453; Conservative 49; Mismatches 8; Indels 4660; Gaps 389;

QY 1 M-----SI-----SS-----D-----E----- 7
Db 1 MGRSPSWLYGVGLGALLLATTCSVNDKNDPTKSSLAFFVFDITGSMFDDLVQREGAAK 60
QY 8 ---V-----N-----F-L-VY----- 13
Db 61 IFKTMAQREKLIYNYIMVPHDPYLGEINTTDTYFMRLSKVYVHGGGDCPEKTLTG 120
QY 14 ---R-Y-----L-QE--SG--F-----SHSAF-T 27
Db 121 ILKALQISLPSSFIYVFTDARSKDYHLEDEVLTIOEKSSVVFVMTGDCGNRTHGFR 180
QY 28 ---FG-I---KS-----H-ISQ-----S-NI----- 39
Db 181 YEKIAAASFGQVHLEKSDVSTVLEVVRHAVKQKVHLMAYEARERGGTVSRNIPVDKHL 240
QY 40 ---N-----GA-----LV-----P-P----- 46
Db 241 ELTISLGDKDDNDLIVLRDPEGRVTKRLYSKEGGTIDLKNVKLIRLKDPSGVWTV 300
QY 47 ---A-----A-----LI-S-I----- 52
Db 301 NNSRLKHTIRVFGHGAVDKGFASRLDRIELARPRVLNODTYLLINMTGLIPPGTV 360
QY 53 --I-----Q-----KGL-----Q-Y--VE--AE- 63
Db 361 GEIDLVDYHSHLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYDEDNIEFMRIAPT 420
QY 64 ---V-----S-I-----N-----ED-----GTLF----- 73
Db 421 AIGSVTVGGPRAFMSPIHQEFVGRDLNCLSTVESASAYTIYWKTEGDIIGGLFVHNTD 480
QY 74 ---D-----G-----R-P-I-----E----- 79
Db 481 TSVWTIPELSLKDAGEYECRVISNNGNYSVKTRVETRESPEPEIFGVRNVSVPLGEAAFLH 540
QY 80 ---S-----L-----SL-I-----DA----- 86
Db 541 CSTRSAGEVEIRWTRYGATVFNPNPTNPTNGTLKIHVTRADAGVYECMARNAGMST 600
QY 87 ---VM-----P-DV-----KLA-----Q-----V- 92
Db 601 RKMRLDIMEPPSVKVTPODVYFNMRBGNLSCEAMGDPKPEVHWYFKGRHLLNDYKYQVG 660
QY 93 Q-----T--R-Q-QAYRD---KLA-----Q----- 105
Db 661 QDSKFLYIRDAATHDEGTYECRAMSQQA-RDITDMLATPPKVEIIQNMVMVGRDRV 719
QY 106 ---Q-----Q-----Q-----A-A-A-A-----A- 111
Db 720 SPECKTIRKPKPKIRWFKNGKDLKPKDDYIKINEQQLHMGAKDEADAGYSCVGENMAG 779
QY 112 ---A-----A-----A-----A-----A----- 114
Db 780 KDQVANLSVGRVPTTIESPHTVRVNIERQVTLQCLAVGIPPEIEWKQGNVLLATLNNP 839
QY 115 ---A-----A-----A-----ASQ----- 120
Db 840 RYQLADGNLLITDAQIEDQGOFTCIARNTYQQSQSSTLMTVGLVSPVLGHVPPBEQLI 899
QY 121 --Q-----GS----- 123

Db 900 EGQDLTSCVVLGTPKPSIWIKDDKPVBEGETIKIEGGSLRLRGNPKDEGKYTCI 959
QY 124 --A--K--N--GE--128
Db 960 AVSPAGNSTLHINVQLIKKPEFVYKPEGGIVFKPTISGMDEKHAVVNSTHVDLGEFA 1019
QY 129 --NT--A-N-G-133
Db 1020 IPCVSGTPPIIITWYLDGRPIITPNSRDFVTADNTLIVRKADKSYGVVTCOATNSAGD 1079
QY 134 -E--AH--TI-A-N--N--144
Db 1080 NEQKTTIRIMNTMISPGQSSFMNVDDLTIPCDVGDGPKPVITWLLDDKPTFEGVNE 1139
QY 137 -G--AH--TI-A-N--N--144
Db 1140 DGLTIPNVNEAHRGFTTCHQAAGNDTRVTLTVHTTPTINAENQEKIALQNDIVLE 1199
QY 145 --HT--D--MME--V--D-152
Db 1200 CPAKALPPVRLWYRGEKIDSQIIPHTIREDGALVQNVKLENTGVFCQVSNLAGEDS 1259
QY 153 --G-DV--EIP--P--NK-AV--163
Db 1260 LSYTLTVHEKPIISEVPGVVDVVGFTIIPCRATGVPFVIRTNWNGIDLQWDEKKFS 1319
QY 164 V--LR--G--H--E--S--170
Db 1320 VDLGLTLRIEADKNDIGNYCVVTVNEAGTSQMTTHVDQEPPIILPSTQNTNNAVVGDR 1379
QY 171 --E--V-F--I-C--175
Db 1380 VELKCYEASPPASVTWFRGIAIGTDTKGYVVEDGLVIQASVEDATYITCKASNPA 1439
QY 176 --A--176
Db 1440 GKAEANLQTVIASPDIKDPDVTOBSIKESHPSLYCPVFSNPLQISWYLNKPLIDD 1499
QY 177 --W--N--178
Db 1500 KTSWKTSDDKRLHVFKAKITDSGVYKCVARNAAGEGSKSFQVEVIVPLNLDSEKYYKKV 1559
QY 179 --PVS--181
Db 1560 FAKGEBVLGCPVSGFPVQINWVDGTVVEPKKYKGATLNDGLTLHFDVSVMQEG 1619
QY 182 --D-L-LA--DS--T--A-R--S 186
Db 1620 NYHCVAQSGNILDIDVELSVLAVPIVGEDDNLVFLGKDISLSCOLQTESDDKTTFVMS 1679
QY 187 --GS--G--T--A-R--194
Db 1680 INGSERPDNVQIIPSDGHLRYITDAKPENNGKYMCRVTNSAGKABRTLTLDVLEPPVFV 1739
QY 195 --IW--N--LS--199
Db 1740 EPVFEANQKIGNPIILOQCVGNPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLTKG 1799
QY 200 --EN--S--T--S-GS--T-207
Db 1800 SAQISCTAENKAGTASRDFIQNTAAPTFRKNEGQETIPRESEITILDCPVLGDFQITW 1859
QY 208 --Q--L-VL--R-H--C--IR 216
Db 1860 MKQGLPLTENDAIPTLNTLRLTILNANRDHDIYTCVANNTAGQVSKDPVWVQVLPKIK 1919
QY 217 --EG--G--Q-DVP--223
Db 1920 NAVVLEINEGEEIILTCDAEGNPTTAKWDFNQGLPKEAVFVNNNHTVVVANNVTKYHT 1979
QY 224 --SNK--DV--T-S-L--D--232

Db 1980 GUVKCVATNKVGOAVKTIINVHVTRKPRFESGLTESELTVNLTRSITECDVDDAIGVGIS 2039
QY 233 W-N--SE-G-TL--L--A-T--GSY--245
Db 2040 WTVNGKPFIAETDGTQTLTAGSRFLHIVSAKTDHDSYACTVTNEAGVATKTNLFVQVPP 2099
QY 246 --DG--F--248
Db 2100 TIVNEGGEVTVIENNSLVLPCEVTGKPNPVVTWKDGRPVGLDKSVQVISEGQOFLVHA 2159
QY 249 --A--R--250
Db 2160 EIAHKSUYICMAKNDVGTAEISFDVDIITRPMIQGIKNIIVTAIKGALPFKPCIPDDKN 2219
QY 251 --IW--T--K-D-G-N--257
Db 2220 FXQIILWLRNYQPIDLEADARITLNSDRRLITILNVTEDEGQYSCRKNDAGENSFDF 2279
QY 258 --L--S-TL--GO-H--265
Db 2280 KATVLVPPITIMLDKDKNTAVEHSTVILSCPATGKPEPDITWFKDGEALHIENIADIIP 2339
QY 266 --K-G-P--I--269
Db 2340 NGELNGNQLKITRIKEGDAGKYTCEDNSAGSVEQDVNVNVITIPKIEKDGPIDYESQ 2399
QY 266 --K-G-P--I--269
Db 2400 NERVISCPVYARPPAKITWLKAGKPLQSDKFVKTSANGQKLYFLKLRDTSKYTCIAT 2459
QY 270 --F--A--L--K--W--274
Db 2460 NEAGTDKRFKVSMLVAPSPDEPNIVRRITVNSGNPSTLHCPAKGSPSTIITWLDGNAI 2519
QY 275 --N--K--276
Db 2520 EPNDRYVFFDAGRQLQISKTEGSDQGYTCIATNSVSGDDELLENTLEVIIPPVIGDERREA 2579
QY 277 --K-G-N--FI--LS--283
Db 2580 VAVIEGSELFCDSNSTGVVEMQKGLTINQDTLURGSFQIOPSSGKMSFLSARKSD 2639
QY 284 --AG--V--DK-T--TI--I--292
Db 2640 SGRYTCIVRNPAEARKLFDFAVNDPPSISDELSSANIQTIVPYVPEINCVVSGSPHPK 2699
QY 293 --W--D-A--H--T--G--298
Db 2700 VYWLFDKPLEPDSAAVELTNNGETLKI VRSQVEHAGTYTCEAQNNGKARKDFLVRVTA 2759
QY 299 --E--AK--Q--302
Db 2760 PHFEKERBEVAVRGDTMLLTCAESSVPLSVYHMHDESQVNGVITSKYAANEKTLN 2819
QY 303 --Q--F--304
Db 2820 VTNIQLDDSEFYCYTAVNEAGITKFFKLI VIETPYFLDQKLYPIILGKRLTDCSATG 2879
QY 305 --P--F--306
Db 2880 TPTPTILFMKGKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSKDMMEV 2939
QY 307 --H--S--308
Db 2940 LLPKLSKWINVEVOAGDPLTLECPEDTSGVHIITWSRQFGKQGLDMRAQSSSKSL 2999
QY 309 --A-P--A--L--V-D--315
Db 3000 YIMQATPEDADSYSCIAVNDAGAEAVFQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059
QY 316 --W--Q--317
Db 3060 EGIPPEIISWFLDGKPILEMPGVYTKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQDT 3119

QY 318 -----S-----N-----N-----T-----F----- 322
Db 3120 YVEISBPRVWASEVMRVVEGRQTIRCEVFGNPEPVNWLKDGPEYTSDLLQFSTKLS 3179
QY 323 -----A-----SC-----S-----TD-----M-----C-----330
Db 3180 YLHLRETTLADGGTYTCIATNKAGESQTTTDEVLVPPRIEDEERVLOQKEGNTVMVHCQ 3239
QY 331 -----I-----HV-----C-----K-----335
Db 3240 VTGRPVVYVTKRNGKEIEQFNPVLRNATRADEGKYSCIASNEAGTAVADFLIDVFTK 3299
QY 336 -----L-----GO-----D-----RP-----I-----K-----343
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRFPFNWDNILLSRPGDTLMILKA 3359
QY 344 -----T-----F-----Q-----346
Db 3360 QRFDGGLYTCVATNSYGDSEQDFKVNVTYKPYIDETIDQTPRAVAGGEIILKCPVLGNPT 3419
QY 347 -----G-----HT-----N-----E-----VN-----A-----I-----355
Db 3420 PVTWKRGDVAVPNSRHTIVNNYDLKINSVTEDAGQYSCIAVNBAGNLTHYAEVIG 3479
QY 356 -----K-----W-----D-----P-----T-----360
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYDRYSISPDGSHITI 3539
QY 361 -----G-----N-----L-----LA-----S-----C-----S-----368
Db 3540 NKAFLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYLECPTS 3599
QY 369 -----D-----M-----LA-----S-----C-----S-----371
Db 3600 GIQPDVIMTKNGMDINMTDSRVLQONNETGNIENVQVTDQRYTCTATNRGKASHDF 3659
QY 372 -----TL-----K-----I-----W-----S-----M-----378
Db 3660 SLDVLSPPFEDIHGTQPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKDSRALDGLTD 3719
QY 379 -----K-----Q-----D-----381
Db 3720 NVDISDGRKLATISQASLENAGLYTCIALNRAGEASLEFKVELSPVIDISKRNDVQPV 3779
QY 382 -----N-----C-----V-----H-----D-----L-----Q-----Q-----H-----390
Db 3780 AVNQPTIMRCVATGHFPFSPISIKWLKNGKEVTDDENIRIVEBQGVQLILRTDSHAGKWSCV 3839
QY 391 -----N-----KB-----I-----Y-----TI-----KW-----399
Db 3840 AENDAGVKELEMLVDVFTPPVSVKSDNPICKALGETITILFCNAGNPYPQLKWAGGSLI 3899
QY 400 -----SP-----T-----G-----P-----GT-----N-----407
Db 3900 FDSPDGARISLAGARLDIHLKKTVDGYTCQALNAAGTSEASVVDVLVPPPEINRDGID 3959
QY 408 -----N-----P-----409
Db 3960 MSPRLPAQOSLTQLCLAQKPVQMRWTLNGTALTHTSTPGITVASDSTFIQINNVSLSDK 4019
QY 410 -----N-----A-----NLM-----LA-----S-----417
Db 4020 GVTCTAENVAGSDNLMYVVDVQAPVINSNGTKQVIEGELAVIECLVEGYPAQVSWLR 4079
QY 418 -----A-----SPD-----421
Db 4080 NGRNVETGVQVRYVTDGRMLTIIERSLDSGIYLCSATNEAGSAQAVTLEVLVSPKII 4139
QY 422 -----ST-----VR-----L-----W-----DV-----429
Db 4140 TSTPGVLTPSSGSKFSLPCAVRGYPDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKABE 4199

RESULT 14
US-10-120-801-75
; Sequence 75, Application US/10120801
; Publication No. US20030203843A1

QY 430 -----D-----R-----431
Db 4200 RHLIVECTAKNDAGADTLEFPVQTIIVAPKISTSGNRYINGSEGTETVIKCEIESSEFS 4259
QY 432 -----G-----I-----C-----434
Db 4260 WSKNGVPLPSNNLIFSEDEYKLIKILSTRLSQGEYSCTAANKAGNATQKTNLNVGVAPK 4319
QY 435 -----H-----TL-----TK-----440
Db 4320 IMERPRTQVVHKGDDQVTLWCEASGVPOPAITWYKONELLTNTGVDETATTKKSVIFSSI 4379
QY 441 -----441
Db 4380 SPSQAGVYTCKAENMVASTEEDIDLIVIMPPEVPERMVSTNPRQTVFLSCNATGIPEP 4439
QY 441 -----H-----Q-----442
Db 4440 VISMRDSNIAIQNNEKYQILGTTLAIRNVLPDDGFYHCIAKSDAGOKIATRKLIYNKP 4499
QY 443 -----E-----P-----V-----Y-----446
Db 4500 SDRPAPIWECDEKPKPKTEYMWDRGDPDNDPQLLPWKDVEDSSLSNGSIAYRCMPGPR 4559
QY 447 -----S-----V-----A-----F-----S-----P-----DG-----454
Db 4560 SSRTVLLHAAPOFIVKPKNTTAAIGAIVELRCSAAGPHPTITWAKDGKLIEDSKFEIAY 4619
QY 455 -----R-----Y-----LA-----S-----G-----460
Db 4680 NOAYSRGLTWEYNGVMPKRLAGIHFMMNGSLVILDTSSLKEGDELELYTCKVNRRRHSI 4739
QY 461 -----S-----F-----DK-----C-----V-----H-----IW-----469
Db 4740 PHLTSAFEGVPEVKTIKVEVNVNGDSVVLDCVTSPLTTHVVTWTKNDQKMLDDDAIYVL 4799
QY 470 -----470
Db 4800 PNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQNLNVYGGSSRRREAYKKENEDASTTIT 4859
QY 470 -----N-----TO-----472
Db 4860 TTSPTTTTETPLTTIIPALITLPAKQYPTDDYHEGSANDDGFPTTQDSUFEPNPLH 4919
QY 473 -----V-----C-----L-----H-----476
Db 4920 PEISVWNTDCAGTINENGDCVDKDKGTHNLKILTGENHCPGEGFAMNPHTRICEDLDECAF 4979
QY 477 Y-----LN-----G-----QVLLN-----LG-----R-----SI-----C-----L-----492
Db 4980 YQCPDFECINYGDFQ-----CNCPGLYELABEGCRDVNECSVRCEGKACFNQLGGYECID 5037
QY 493 -----YTL-----P-----H-----H-----L-----V-----500
Db 5038 DPCPANYSLVDDRYCEPECENCTSTPIQVHMLAIPSGLPISHIATLTAYDKSGRVLNDTT 5097
QY 501 -----VI-----PL-----V-----A-----L-----I-----EL-----510
Db 5098 YAISTDGAPLARGMTSGPFTIKAVKRGHAQVWTRVLAAGDHHKVRVRAHSDHATNELH 5157
QY 511 -----LVL-----513
Db 5158 APKETNFLVL 5167

GENERAL INFORMATION:
APPLICANT: Pena, Carol
APPLICANT: Guo, Xiaojia
APPLICANT: Shimkets, Richard
APPLICANT: Padigar, Muralidhara
APPLICANT: Kekuda, Ramesh
APPLICANT: Spytek, Kimberly
APPLICANT: Mehraban, Fuad
APPLICANT: Topper, James N.
APPLICANT: Malyankar, Uriel
APPLICANT: Wasserman, Scott
APPLICANT: Edinger, Shlomit
APPLICANT: Smithson, Glennda
APPLICANT: Gunther, Erik
APPLICANT: Komuves, Laszlo
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-340
CURRENT APPLICATION NUMBER: US/10/120,801
CURRENT FILING DATE: 2002-04-11
PRIORITY APPLICATION NUMBER: 60/285748
PRIORITY FILING DATE: 2001-04-23
PRIORITY APPLICATION NUMBER: 60/286068
PRIORITY FILING DATE: 2001-04-24
PRIORITY APPLICATION NUMBER: 60/286292
PRIORITY FILING DATE: 2001-04-25
PRIORITY APPLICATION NUMBER: 60/288334
PRIORITY FILING DATE: 2001-05-03
PRIORITY APPLICATION NUMBER: 60/291241
PRIORITY FILING DATE: 2001-05-16
PRIORITY APPLICATION NUMBER: 60/322284
PRIORITY FILING DATE: 2001-09-14
PRIORITY APPLICATION NUMBER: 60/285609
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 75
LENGTH: 5198
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-10-120-801-75

Query Match 70.9%; Score 2586; DB 14; Length 5198;
Best Local Similarity 9.9%; Pred. No. 2.8e-29;
Matches 423; Conservative 70; Mismatches 19; Indels 3770; Gaps 370;

QY 1 M-----SI-----SS-----D-----E-----7
DB 1 MGRSPSWLYGVLLALLATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLVQVREGAAK 60
QY 8 -V-----N-----F-----L-----VY-----13
DB 61 IFKTVMAQREKLIYINMVPFHPDYLGEINTDSTYFMRQLSKVYVHGGGDCPEKTLTG 120
QY 14 -----R-----Y-----L-----QE-----SG-----F-----SHSAF-T 27
DB 121 ILKALQISLPSSFIYVFTDARSKDYLEDEVLNTIQEKSSVVFVMTGCGNRTHPGFRT 180
QY 28 -----PG-I-----KS-----H-ISQ-----S-NI-----39
DB 181 YEKIAAASFQVPHLEKSDVSTVLEYVRHAVKQKVHLMYAEARGGTVSRNIPVDKHL 340
QY 40 -----N-----N-----GA-----LV-----P-P-----46
DB 241 ELTISLSDKDDSDNLDIVLRDPDEGTVDKRLYSKEGGTIDLKVKLIRLKDPSGQVTV 300
QY 47 -----A-----A-----A-----LI-----S-I-----52
DB 301 NTSRLKHTIRVFGHGAVDKYGAFASRLDRIELARPRVNLQDTYLLINMTGLIPPGTV 360
QY 53 -I-----Q-----KGL-----Q-Y-----VE-----AE-63
DB 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPFVPPKGLFFVRVQYGEDNDFEMRIAPT 420

QY 64 -V-----S-I-----N-----ED-----GTLF-----73
DB 421 AIGSVIVGGPRAFMSPIHQEFVGRDLNLSCTVESASAYTIYWKGTGEDIIIGPLFVHN 480
QY 74 -----D-----G-----R-----P-I-----E-----79
DB 481 TSVWTIPELSLKDAGEYECRVISNNGNSVSKTRVETRESPPFIQVVRNVSPVIGEAFLH 540
QY 80 -S-----L-----SL-I-----DA-----86
DB 541 CSTRSAGEVEIRWTRYGATVFNQPNTERNTNGTLKIHVTRADAGVYECMARNAGCMST 600
QY 87 -VM-----P-DV-----V-----92
DB 601 RKMRLDIMEPPSVKVTQDVYFNMRGVNLSCEAMGDKPVEVHWYFKGRHLLNDYKQVG 660
QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q-----105
DB 661 QDSKFLYIRDAATHDEGTVECRAMSQAQA-RD-TDLMLATPPKVEIIQNMVMVGRDRV 719
QY 106 -----Q-----Q-----A-----AAA-----AA-113
DB 720 SPECTIRKPHPKIRWFKNGKDLIKPDDYIKINEGQLHIMGAKDEADAGAYSCVGNMAG 779
QY 114 -AAAA-----AS-----Q-----Q-----GS-----A-KN-126
DB 780 KDQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNP 839
QY 127 -GEN-----T-----A-N-----132
DB 840 RYTQLADG-NLLITDAQIBDOGQFTCIARNTYGOOSQSTTLMVTLGLVSPVGHVPEEQ 898
QY 133 -----G-----EE-----N-----G-----137
DB 899 IEGQDLTLCVVVLGTPPKPSIWIKDDKPVESGPTIKIEGGSGLLRLRGNPKDEKGYTC 958
QY 138 -A-----H-----TI-----A-N-NH-----145
DB 959 IAVSPAGNSTLHINVQLIKKPFVYKPEGGIKPTISGMDEKHVAVVNSTHVDLDEGF 1018
QY 146 -----T-----146
DB 1019 AIPCVVSGTPPIITWYLDGRBITPNSRDTVTADNTLIVRKADKSYSGVYTCQATNSAG 1078
QY 147 D-----M-----M-----E-----VD-----GD-----V-155
DB 1079 DNEQKTTIRIMTPMISPGSSFNMMVVDLFTIPCDVYGDPKPVITWLLDDKPFTEGVN 1138
QY 156 E-----IP-----P-----N-K-A-----VVL 165
DB 1139 EDGSLTIPNVNEAHRGFTTCHAAQNAAGNDTRVTTLTVHTTPTINAENQEKIALQNDIVL 1198
QY 166 -----R-----G-----167
DB 1199 ECPAKALPPPRLMTVEGEKIDSQLIPHTIREDGALVQNVKLENTGVFVQVSNLAGE 1258
QY 168 -----HE-----SEV-----F-----I-C-A-----WNP-----V-180
DB 1259 SLUSYTLTHHEKPKIIESEVPGVVDVVGKFTIEIPCRATGVPEVIRTNKNGIDLKMEKKF 1318
QY 181 S-D-L-L-----AS-----G-----SG-----D-----S-----TA-----193
DB 1319 SVDNGLTLRIYEAADKNDIGNYCVVNEAGTSQMTTHVDVQBPPIILPSTQTNNTAVVGD 1378
QY 194 RI-----W-----NL-----S-ENST-----S-204
DB 1379 RVELKCYEASPPASVTFWFRGIAIGTDTKGYVESDGLTVIQASVEDATIYTCASNP 1438
QY 205 -G-----S-----TO-----L-----V-----L-211
DB 1439 AGKAENLOVTVIASPDIKDPDVVTQESIKESHPSLYCFVSNPLPQISWYNLNDPLID 1498
QY 212 -----R-H-----CI-R-----216

Db 1499 DKTSWKTDDKRLHVFKAKITDSGVYKCVARNAAGEGSKSFQVEVIVPLNLDESQYKXK 1558
QY 217 --EG--G--Q--D--V--P--SN--225
Db 1559 VFAKEEVEVTLGCPVSGFPVPQINWVVDGTVBPGKKYKATLSDGLTLHFDVSQVKQE 1618
QY 226 --N--SE--K--DVT--S--L--D--W--233
Db 1619 GNVHCAQSKGNLIDVELSVLAVPIVGEDDNLVFLGKDISLSCDLQTESDDKTTFFW 1678
QY 234 --N--SE--G--TL--L--240
Db 1679 SINGESDRPNVQIIPSDGRLYITDAKPNNGKYMCRVNTSAGKAERTILTLDVLEPPVF 1738
QY 241 --A--TG--SY--D--246
Db 1739 VEPVFANOKLIGNPILLOQVGTGNPKPTVIWKIDGNDVDSWLFDESLSLLRIEKLGT 1798
QY 247 --G--F--A--R--I--251
Db 1799 KSAQISCTAENKAGTASRFFIQNIAAPTFKNEGDOETIPRESEITITLDCPVSLGFOIT 1858
QY 252 W--G--T--KD--255
Db 1859 WMKQGLPLTENDAIFTLDNTRLTILNANRDHDIYTCVANNTAGQVSKDFVVVVQVLPKI 1918
QY 256 --CN--L--A--S--260
Db 1919 KNAVTLNEGEBEIILTCAEGNPPTTAKWDFNOGDLPEKAVFVNNHTVVVNNVTKYH 1978
QY 261 T--L--GO--H--K--266
Db 1979 TGVYKYATNKVQAVKTIHVHRTKRPESGLTESELTVNLTRSLTLECDVDDAIGVGI 2038
QY 267 --G--P--I--F--270
Db 2039 SWTVNGKPLAETDGVOTLAGRFLHVSAKTDDHGSYACTVNEAGVATKTNLFPQVP 2098
QY 271 --AL--K--W--274
Db 2099 PTIVNEGBYTVIENNSLVLPCEVTGKPNPVTWKDGRPVGLKSVQVLSQGPVKIVH 2158
QY 275 --N--K--KG--278
Db 2159 AETAHKSVCMAKNDVGTAEISFDVDIITRPMIQIKNIIVTAIKGGALPKFCPIDDDK 2218
QY 279 NF--I--LS--AG--285
Db 2219 NFKGQIILWRNYQPIDLEAEDARIITLSNDRRLTILNVTEDEBGQYSCRVKNDAGNSFD 2278
QY 286 --V--D--KT--TI--I--W--D--A--H--296
Db 2279 FKATVLVPTTIIMLDRKDKTAVEHSTVTLSCPATGKPEPDIITWPKDGEAIIHENIADII 2338
QY 297 --T--G--BA--K--Q--302
Db 2339 PNGELNGNQLKITRIKEGDAGKTCADNSAGSVEQDVNNVITIPKIEKGIPSDYESQ 2398
QY 303 Q--F--P--H--303
Db 2399 QNERVVISCPVYARPPAKITWLKAGKPLQSDKFVKTSAHQKLYLFLKRETDSKTYCIA 2458
QY 304 --F--P--H--307
Db 2459 TNEAGTKRDFKVMVLVAFSFBEPNIVRITVNSGPNSTLHCPAKGSPSTITWLKDGNA 2518
QY 308 --S--A--P--310
Db 2519 IEPNDRYFDFAGRQIQISKTEGSDQGYTCATNSVSGDDLENTLEVIPPVIDGERRE 2578
QY 311 --A--L--D--VD--WQ--S--N--NT--F--A--323

Db 2579 AVAVIEGFSSELFCDNSNSTGVDVEMQKQGLTINQDTRLGRDSFIQIPSSGKKMSFLSARKS 2638
QY 324 --S--C--S--TD--M--CI--H--332
Db 2639 DSGRYTCIIVRNPAEGARKLDFFAVNDPPSISDSSANIQTIVPYVPVEINCVSQSGPH 2698
QY 333 --V--CK--LG--Q--D--339
Db 2699 KVTWLFDDKPLEPDSAAVELTNGNETLKIVRSQVEHAGTYTCEAQNNGVKARKDFLVRVT 2758
QY 340 --R--P--I--KT--344
Db 2759 APPHFEKEREVVARVGDTMLLTCAESSVPLSVYWHADSVQNGVITSKYAANEKTL 2818
QY 345 --F--O--346
Db 2819 NVTNIQLDEGFFYCTAVNEAGITKFFKLVITETPYFLDQQLYPIILGKRLTLDCSAT 2878
QY 347 G--H--T--N--EV--N--A--354
Db 2879 GTPPTILFMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGGRYTCIAENKAGRSEKMMVE 2938
QY 355 I--K--W--DP--T--G--N--362
Db 2939 VLLPPKLSKEWINVEYQAGDPLTLECPIEDTSCVHITWSRQFGKQGLDMRAQSSDKSK 2998
QY 363 LL--A--SC--S--D--DMTKI--375
Db 2999 LYIMQATPEDADSYSCIAVNDAGGABAVQVTVNTPPKIFGDSFSFSTEIVADTTLIPCR 3058
QY 376 --W--SM--KO--DN--CV--384
Db 3059 TEGIPPEISWFLDGKPILEMPGVYKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQD 3118
QY 385 --384
Db 3119 TVVEISEPPRVVMASEVMRVEGRQTTIRCEVFGNPEPVVNMKDGEPYTSDDLQFSTKL 3178
QY 385 --H--D--L--Q--O--H--390
Db 3179 SYLHRETTLAGGTYTCIATNKAQESQTTTDEVLVPPRIEDERVLQGEKNTYVMVHC 3238
QY 391 --N--KEI--Y--395
Db 3239 QVTGRPVVYWKNGKEIEQFNPVLHINATRADEGKISCIASNEAGTAVADFLIDVFT 3298
QY 396 --T--I--K--W--SP--401
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISMLKGGRPFNMDNIILSPRGDTLMILK 3358
QY 402 --T--3403
Db 3359 AORPDGLYTCVATNSYGDSEQDFKVNVTYKPIDETIDQTPKAVAGSEIILKCPVLGNP 3418
QY 404 --P--G--T--NN--408
Db 3419 TTVTWKRGDDAVPNDPSRHTIYNNVDLKINSVTTEDAGQYSCIYVNEAGNLTHYAAEVI 3478
QY 409 --P--N--AN--L--M--LASAS--F--420
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYLDVYISPDGSHIT 3538
QY 421 --D--S--T--V--R--L--426
Db 3539 INKAKLSGKGYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYECPI 3598
QY 427 --W--DV--D--R--GI--CI--H--436
Db 3599 SGIPQPDVITWTKNGMDINNTDSRVILAQNNEFTFGIENVQVTDQGYTCATNREGKASHD 3658
QY 437 --TIT--K--H--Q--E--P--444
Db 3659 FSLDVLSPPEFDIHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALODLT 3718

QY 445 --V--YS-VA-----F--SP-----D-----453
Db 3719 DNVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPVIDISRDNDVQ PQ 3778
QY 454 -----G-----R-----Y-L-AS--G--S-461
Db 3779 VAVNQPTIMRCAVTHGHPFSPKWLKNGKEVTDENIRIVEQGQVLQILRTDSDHACKWSC 3838
QY 462 -----F-----D--K-----C-----465
Db 3839 VAENDAGVKELEMDVFTPPVSVKSDNPIKALGETITLFCNASGNPPQLKWAAGGSL 3898
QY 466 -----V-H-----I-----468
Db 3899 IFDSPGARISLKGARLDIPLHKKTDVGYTCOALNAAGTSEASVSDVLPPEINRDI 3958
QY 469 -----W-N-----T-Q-----472
Db 3959 DMSPLPQAQSLTLQCLAQKPPQMRWTGLNGTALHSTPGITVASDSTFIQINNVSLS 4018
QY 473 --V--C-----LHY-----L-NG--QV-----L-484
Db 4019 KGYTCTAENAVAGSDNLMNVVQAPVISNGGTQKQVIEGELAVIECLVEGYPAQVSWL 4078
QY 485 --N-----L--GR-----S-I--C-----LVTL-----495
Db 4079 RGNRVETGVQVRYVTDGRMLTIIERSLDSGIYLCSATNEAGSAQAYTLEVLSPKI 4138
QY 496 -----P-----H-----497
Db 4139 ITSTPGVLTSSGSKFSLPCAARGYPDPPIISWTLNGNDIKDNGHTIGADGTLHIEKAE 4198
QY 498 --HLV-----V--I--P-----LV-----AL--IE-----509
Db 4199 ERHLIYECTAKNDAGADTLEFPVQTTIVAPKISTSGNRYNGSEGTETVIKCEIESSEB 4258
QY 510 -----L--LV-----LK 514
Db 4259 SWSKNGVPLPSNNLIFSEDYK 4280

RESULT 15

US-10-369-493-6858
; Sequence 6858, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6858
; LENGTH: 5198
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6858

Query Match 70.9%; Score 2586; DB 14; Length 5198;
Best Local Similarity 9.9%; Pred. No. 2.8e-29;
Matches 423; Conservative 70; Mismatches 19; Indels 3770; Gaps 370;
QY 1 M-----SI-----SS-----D-----E-----7
Db 1 MGRSPSWLYGVLGILLIATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLVOVREGAAK 60

QY 8 --V--N-----F--L--VY-----13
Db 61 IFKTVWAQREKLIYNYIMVPHDPYLGEIINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120
QY 14 -----R--Y-----L--OE--SG--F-----SHSAP--T 27
Db 121 ILKALQISLPSSFYVFTDARSKDYHLEDEVLTNIQEKQSSVVFVMTGDCGNRTHGPFRT 180
QY 28 -----RG-I--KS-----H-ISO-----S-NI-----39
Db 181 YEKIAAASFGQVHLEKSDVSTVLEYVRHAVKQKKVHLMYERERGGTVSRNIPVDKHL 240
QY 40 -----N-----GA-----LV-----P-P-----46
Db 241 ELTISLSDGKDDNDLIVLRDPEGTVDKRLYSKEGGTIDLKNVKILRLKPSPGVWTV 300
QY 47 -----A-----A-----LI--S--I-----52
Db 301 NTNSRLKHTIRVFGHGAVDKYGFAFRPLDRIELARPRVLNQDTYLLINMTGLIPPGTV 360
QY 53 --I-----Q-----KGI-----Q-Y--VE--AE-63
Db 361 GBIDLVDYHGHSLYKAVASPHRTNPNMYPAGFPVPPKGLFFVRVQSDYEDNVEFMRIAPT 420
QY 64 --V--S--I-----N-----ED--GTLF-----73
Db 421 AIGSVLVGPRAPFPIHQEFVGRDLNLSCTVESASAYTIYVWKTGEDIIIGPLFYHNTD 480
QY 74 -----D-----G-----R--P-I-----E-----79
Db 481 TSVWITIPELSLKDAGEYECRVISNNGVSVKTRVETRESPPEIFGVNRVSVPLGEAFLH 540
QY 80 --S-----L-----SL-I-----DA-----86
Db 541 CSTRSAGEVEIRWTRYGATVFNQPNPTNPTNGTLKHHVTRADAGVCECMARNAGMST 600
QY 87 --VM-----P-DV-----V-----92
Db 601 RKWRDLIMEPPSPVKVTPQDVYFNRMEGVNLSCEAMGDKPEVHWYFKGRHLLNDYKQVG 660
QY 93 Q-----T--R--Q--QAYRD-----KLA-----Q-----105
Db 661 QDSKFLYIRDAATHDEGTCECRAMSQAQA-RDITDLMLATPPKVEIIONKMMVGRGDRV 719
QY 106 -----Q-----Q-----Q--A--AAA-----AA-113
Db 720 SPECKTIRKPKPKIRWFKNGKDLKPDYIKINEGQLHMGAKDEDAAGAYSCVGENMAG 779
QY 114 --AAAA-----AS-----Q--Q-----GS--A--KN-126
Db 780 KDQVANLSVGRVPPTIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNP 839
QY 127 -----GEN-----T--A--N-----132
Db 840 RVTQLADG-NLLITDAQIEDQGTCTIARNTYQSQSQTLLMVTGLVSPVLGHVPPEEQ 898
QY 133 -----G-----EE-----N--G-----137
Db 899 IEGQDLTSLCVVLTGPKPSIWMKDDKPVESGPTIKIEGGSLRLRGNPKDEKGYTC 958
QY 138 --A-----H-----TI-----A--N--NH-----145
Db 959 IAVSPAGNSTLHNVQLIKKPEVYKPEGGIVPKPTISGMDEKHVAVNSTHVDLGE 1018
QY 146 -----M--M--E-----VD-----GD-----T-----146
Db 1019 AIPCVSVSGTPPPIITWYLDGRPIITNSRDTVTADNTLIVRKADKSYSGVYTQATNSAG 1078
QY 147 D-----M--M--E-----VD-----GD-----V-155
Db 1079 DNEQKTTIRIMNTPMISPGQSSFNWVVDLFTIPCVDYGDPKPVITWLLDDKPFTEGVN 1138

QY 156 E-----IP-----P-----N-K-A-----VVL 165
 Db 1139 EDGSLATIPVNEAHRGFTTCHQAQAAGNDRTTLTVHTTPTINAENOEKIALQNDQIVL 1198
 QY 166 -----R-----G----- 167
 Db 1199 ECPAKALPPVRLWYVEGEKIDSLIPHTIREDCALVLQNVKLENTGVFCVQVSNLAGED 1258
 QY 168 -----HE-----F-----I-C-A-----WNP-V----- 180
 Db 1259 SLSYTLTVHEKPKLISEVPGVVDVVGFTTIEIPCRATGVPEVIRTNWKNIGIDUKMEKKF 1318
 QY 181 S-D-L-L-----AS-----G-----SG-----D-----S-----TA----- 193
 Db 1319 SVDNLGTLRIYADKNDIGNYCNVWNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGD 1378
 QY 194 RI-----W-----NL-----S-ENST-----S----- 204
 Db 1379 RVELKCYVEASPPASVTWFRFGIAIGTDTKGYVVESDGTLVIOQASVEDATIYCKASNP 1438
 QY 205 -G-----S-----TO-----L-V-----L----- 211
 Db 1439 AGKAEBANLQVTIASPDIDKDPDVVTQESIKESHPPFSLYCPVFSNPLPQISWYLNKPLID 1498
 QY 212 -----R-H-----CI-R----- 216
 Db 1499 DKTSWKTSDDKRKLHVFKAKITDSGVYKCVARNAAGBSKSFQVEVIVPLNLDESXYKKK 1558
 QY 217 -EG-----G-----Q-----D-----V-P-----SN----- 225
 Db 1559 VFAKEGEVTLGCPVSGFPVQPNWVVDGTVVFBPKKYGATLSNDGLTLHFDSVSVKOE 1618
 QY 226 -----KVT-S-L-----D-----W 233
 Db 1619 GNHVCVAQSGKNILDIDVELSVLAVPIVGEDDNLVFLGKDLSLSCDLQTESDKKTTFW 1678
 QY 234 -N-SE-----G-----TL-L----- 240
 Db 1679 SINGSEDRPDNVQIPSDGHRLYITDAKENNGKYMCRVNSAGKAERTLTLDVLEPPVF 1738
 QY 241 -----A-----TG-----SY-D----- 246
 Db 1739 VEPVFEANOKLGNPILLOQVGTGNPKPTVIWKIDGNDVDSKSWLFDLSLLRIEKLIG 1798
 QY 247 -----G-----F-----A-----R-----I- 251
 Db 1799 KSAQISCTAENKAGTASRDPFQIONIAAPTKEGQDFTIPRESEITITLDCPVSLGDFQIT 1858
 QY 252 W-----T-----KD----- 255
 Db 1859 WMKQGLPLTENDALFTLDNTRTLILNANRDHDIYTCVANNTAGQVSKDFVQVQVLPKI 1918
 QY 256 -----GN-----L-A-----S----- 260
 Db 1919 KNAVVTLEINEGEIILTCDAENPTPTAKWDFNQGDLPKEAFVANNHTVWVNNVTKYH 1978
 QY 261 T-----L-----CO-----H-K----- 266
 Db 1979 TGVIKCYATNKVGQAVKTNVHVRTKPRFESGLTESELVNLFRSITLCECDVDDAIGVGI 2038
 QY 267 -----G-P-----I-----F----- 270
 Db 2039 SWTVNGKPLAETDGVQTLAGGRPLHIVSAKTDHGSYACTVTNEAGVATKTNLFVQVP 2098
 QY 271 -----AL-----K-----W----- 274
 Db 2099 PTIVNEGGEVTVIENNSLVLPCVETGKPNPVVWTKDGRPVGDLKSVQVLSQEQFKIVH 2158
 QY 275 -----N-----K-----KG----- 278
 Db 2159 AEIAHRGSYICMAKNDVGTAEISFDVDTIIPRMIQRIKNIIVTAIRGGALPFXCPIDDDK 2218
 QY 279 NF-----I-----LS-----AG----- 285

Db 2219 NFKGQIILWRNYQPIDLEAEADARITRLSNDRRRLTILNVTENDEQYQSCRKNDAGENSFD 2278
 QY 286 -----V-----D-KT-----TI-----I-W-D-A-H----- 296
 Db 2279 FKATVLVPPPTIIMLDKDKNKAVEHSTVTLSCPATKPKPEPDIITWFKDGEAIIHENIADI 2338
 QY 297 -----T-----G-----EA-----K-----Q 302
 Db 2339 PNGELNGNLKTRIKEGDAGKYTCADNSAGSVEQDVNVNVTIPKIEKDGPSPDYEQ 2398
 QY 303 Q----- 303
 Db 2399 QNERVVISCPVAVRPPAKITWLKAGKPLQSKFVKTSANGQKLYLFKLRETDSSKYTCIA 2458
 QY 304 -----F-----P-P-----H----- 307
 Db 2459 TNEAGTKRDFKVKVSMVLVAFSFBEPNIVRTVNSGNPSTLHCPAKGSPSPITWLKDGNA 2518
 QY 308 -----S-----A-----P----- 310
 Db 2519 IEBNDRYVFDAGRQLOISKTEGSDOGRYTCIATNSVSGDDLENTLEVIIPPIDGERRE 2578
 QY 311 -A-----L-D-----VD-WQ-----S-N-NT-----F-A----- 323
 Db 2579 AVAVIEGFSSELFCDSNSTGVDVWQKGLTINQDTLRGDSFTQIPSSGKMSFLSARKS 2638
 QY 324 -S-----C-----S-TD-M-----CI-----H- 332
 Db 2639 DSGRYTCIVRNPAGEARKLFDFAVNDPPPSIDELSSANITQIVPYVPVINCINVSUSGPH 2698
 QY 333 -----V-----CK-----LG-O-D----- 339
 Db 2699 KYWLFDDKPLPDSAAELTNNGETLKIIVRSQVEHAGTYTCEAONNVGKARKDFLRVT 2758
 QY 340 -----R-----P-----I-----KT- 344
 Db 2759 APHPFEKEREVVARVGDWMLTLCNAESSVPLSVVYHHAHDESQVGVITSKYAANEKTL 2818
 QY 345 -----F-O----- 346
 Db 2819 NVTNIQLDDGYFYTCVAVNEAGITKFKPLIVETPYFLDQOKLYPIILGKRLTLDOSAT 2878
 QY 347 G-----H-----T-N-----EV-----N-A----- 354
 Db 2879 GTPPPITLFWKDKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVE 2938
 QY 355 I-----K-W-----DP-----T-----G-----N 362
 Db 2939 VLLPPKLSKEWINVEVQAGDPLTLECPIEDTSGVHITWSRQFGKQGLDMRAQSSSDSK 2998
 QY 363 LL-----A-SC-----S-----D-----DMTLKI----- 375
 Db 2999 LXIMOATPEDADSYSCIAVNDAGBAEVQVTVNTPPKIFGDSFSTTEIVADTTLEIPCR 3058
 QY 376 -----W-----SM-----KO-----DN-----CV----- 384
 Db 3059 TEGIPPEBSWFLDGPILPMPGVYKQGLSLRIDNIPKQEGRYTCVAKNAGRAEQD 3118
 QY 385 ----- 384
 Db 3119 TYVEISEPPRVVMASEVMRVVEGRQTTIRCEVFNPEPVVNLKDGEPYTSLLQFSTKL 3178
 QY 385 H-----D-L-----Q-----O-----H- 390
 Db 3179 SVLHRETTLADGGTYTCIATNKAGESOTTDDVEVLPPRIEDEERVLOGEGNTYMHVC 3238
 QY 391 -----N-KEI-----Y----- 395
 Db 3239 QVTGRPVVVTWKRNGKEIEQPNVLHIRNATRADEGKYSCIASNAGTAVADFLIDVFT 3298
 QY 396 -----T-----I-----K-----W-----SP----- 401

Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGGRPFNMDNIILSPRGDTLMILK 3358
QY 402 -----T-----G-- 403
Db 3359 AQRFDGLYTCVATNSYGDSEQDFKVNVTYKPYIDETIDQTKRAVAGEIILKCPVLGNP 3418
QY 404 -P-----G-----T--NN----- 408
Db 3419 TPTVTWKRGDADVNDPSRHTIVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVI 3478
QY 409 --P-----N--L-M-----LASAS-F----- 420
Db 3479 GKPTFVRKGNLVEIENDTITWDCGVTSRPLPSISWFRGDKPVLYDRYSISPDGSHIT 3538
QY 421 -----D-----S-----T-----V-R--L----- 426
Db 3539 INKALSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPI 3598
QY 427 -----W-----DV--D-R-----GI-----CI-----H- 436
Db 3599 SGIPQPDVITWKNGBDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNREGKASHD 3658
QY 437 ----TLT--K--H--Q-----E-----P----- 444
Db 3659 FSLDVLSPPEFDIHGTQPTIKREGDTITLTCPKLAEDIAQVMDVSWTKDSRALDGLT 3718
QY 445 --V-----YS-VA-----F-----SP-----D----- 453
Db 3719 DNVDISDGRKLTISOASLENAGLYTCIALNRAGEASLEFKVEILSPVVIDISRNDVQFQ 3778
QY 454 -----G-----R-----Y-L--AS--G-S- 461
Db 3779 VAVNQPTIMRCAVTHPPFSIKWLKNGKEVTDENIRIVEQGVQLILRTDSDHAGKWSC 3838
QY 462 -----F-----D--K-----C----- 465
Db 3839 VAENDAGVKELEMLVDFTPPVSVKSDNFIKALGETITLFCNASGNPYPQLKWKAGGSL 3898
QY 466 -----V-H-----I 468
Db 3899 IPDSPGARISLKGARLDIPLHKKTDVGDVTCQALNAAGTSEASVSDVLVPPPEINRDI 3958
QY 469 -----W--N-----T--Q----- 472
Db 3959 DMSPLPAQOSLTLOCLAQKPVQMRWTLNGTALTHTSPGITVASDSTFIQINNVSLS 4018
QY 473 --V-C-----LHY-----L-NG--QV-----L 484
Db 4019 KGVYTCVAENVAGSDNLMYNDVVQAPVISNGGKQVIEGELAVIECLVEGYPAPQVSWL 4078
QY 485 ---N-----L--GR-----S-I-C-----LYTL----- 495
Db 4079 RNGNRVETGVQVRYVTDGRMLTIIEARSLDSGILCSATNEAGSAQAYTLEVLVSPKI 4138
QY 496 -----P-----H----- 497
Db 4139 ITSTPGVLTPSSGSKFSLPCAVRGYDPPIISWTINGNDIKDGENGTIGADGTLHIEKAE 4198
QY 498 --HLV-----V--I--P-----LV-----AL-----IE----- 509
Db 4199 ERHLIVECTAKNADAGATLEFPVQTIIVAPKISTSGNRYNGSEGTETVIKCEIESESEF 4258
QY 510 -----L-----LV-----LK 514
Db 4259 SWSKNGVPLLPNSNLIIFSEDIK 4280

Search completed: January 3, 2005, 15:48:12
Job time : 137 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:09:41 ; Search time 21.6667 Seconds
(without alignments)
2282.558 Million cell updates/sec

Title: US-09-987-701-4

Perfect score: 3649

Sequence: 1 MSISDEVNFLVRYLQESG.....LPHLVVIPLVALIELLVLK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2589.3	71.0	7962	I38346	elastic titin - hu
2	2588	70.9	5175	T20992	hypothetical prote
3	2586	70.9	5198	T43290	hemacentin precurs
4	2585	70.8	10797	T30192	probable peptide s
5	2575.8	70.6	26926	I38344	titin, cardiac mus
6	2575.1	70.6	4660	T42737	gp330 protein prec
7	2574.2	70.5	7829	T15789	hypothetical prote
8	2572.1	70.5	6658	T13931	projectin - fruit
9	2568.8	70.4	5232	A45086	HC-toxin synthetas
10	2568.3	70.4	3461	S58870	reelin precursor -
11	2566.1	70.3	5825	T12117	polyprotein - fava
12	2562.6	70.2	4544	S02392	alpha-2-macroglobu
13	2556	70.0	4543	A53102	alpha-2-macroglobu
14	2553.7	70.0	4753	A47437	LDL-receptor-relat
15	2552.7	70.0	5376	T42215	zonadhesin - mouse
16	2551.4	69.9	4545	S25111	alpha-2-macroglobu
17	2550.6	69.9	6831	A88852	protein unc-22 (im
18	2550.6	69.9	6839	S57242	twitchin (similar
19	2550.6	69.9	7160	T27935	hypothetical prote
20	2549.2	69.9	8243	T31307	type I fatty acid
21	2549.1	69.9	6805	S20901	titin - rabbit (fr
22	2548.4	69.8	6669	S55024	nebulin, skeletal
23	2537.3	69.5	9376	T14593	syringomycin synth
24	2533.4	69.4	15281	S41309	cytosporin synth
25	2528.8	69.3	3623	T09456	intrinsic factor-B
26	2527.7	69.3	4861	S71752	giant protein p619
27	2527.3	69.3	4767	T31345	hypothetical prote
28	2527.1	69.3	3623	T08618	intrinsic factor-B
29	2527	69.3	4836	T14346	herc2 protein - mo

30	2523	69.1	5369	2	T44807	myosubtilin synth
31	2516.8	69.0	4464	2	D87755	protein T21E12.4 [
32	2516.6	69.0	5138	2	B96695	hypothetical prote
33	2515.4	68.9	4302	2	A38971	polycystic kidney
34	2514.8	68.9	4930	2	E69679	polyketide synthet
35	2514.5	68.9	3871	2	T22812	hypothetical prote
36	2513.4	68.9	4568	2	T08030	dynein beta heavy
37	2511.8	68.8	4572	2	S57908	hypothetical 537K
38	2502.3	68.6	4924	2	T50176	probable peptide s
39	2501.8	68.6	6359	2	T31679	bacitracin synthet
40	2500.9	68.5	4644	1	A38905	dynein heavy chain
41	2500.5	68.5	4447	2	A69679	polyketide synthas
42	2498.9	68.5	4639	1	A54794	dynein heavy chain
43	2498.8	68.5	6642	2	T29757	protein UNC-89 - C
44	2496.5	68.4	6486	2	T31076	tyrocidine synthet
45	2496	68.4	3856	2	T51174	ataxia-telangiecta

ALIGNMENTS

RESULT 1

I38346

elastic titin - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004

C:Accession: I38346

R:Label: S.; Kolmerer, B.

Science 270, 293-296, 1995

A:Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.

A:Reference number: A57430; MUID:96026330; PMID:7569978

A:Accession: I38346

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-7962 <RES>

A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g101

C:Genetics:

A:Gene: GDB:TTN

A:Cross-references: GDB:127867; OMIM:188840

A:Map position: 2q31-2q31

Query Match 71.0%; Score 2589.3; DB 2; Length 7962;
Best Local Similarity 8.8%; Pred. No. 1.3e-52;
Matches 454; Conservative 50; Mismatches 8; Indels 4667; Gaps 390;

QY	1	MS-----IS-----SDE-----V-N-----	-----9
DB	57	MTQFEDTYQLEIAEAYPEDEGTYTFVANNVQVSTANLSLEAPESILHERIEQIEME	116
QY	10	-----F-----L-----VY-----R--Y--15	
DB	117	MKAAPVIRKRIEPLVALGHLAKFTCEIQSAPNRFQWFKAGREIYSDKCSIRSSKYIS	176
QY	16	---L--Q-----E-----S-----G---F--21	
DB	177	SLEILRTQVDCGEYTCRASVEYGVSCATLTIVTEAYPTLSRPKSLTTFVGKAKFI	236
QY	22	-----S-----H-----	24
DB	237	CTVTGTPVETIWKDGAALSPNWRISDAENKHLELSNLTIDRGVYVSCSKNFKGA	296
QY	25	---A-----F-----T-----	27
DB	297	DTQCAELIIDKPHFIKELEPVQSAINKKVHLEQCVDKRVTVTWMSKDGQKLPPGKDYK	356
QY	28	--F-----G-----IK-----	31
DB	357	ICFEDIATLEIPLAKDKSGTYVCTASNEAGSSCSATVTVREPPSFVKVYDPSYLMLP	416
QY	32	--S---H-----ISQ-----SN-----IN-----G-----41	
DB	417	GESARLHCKLKGSPVI-QVTFPKNNKELSESNTVRMYFVNSEAILDITDKVEDSGSYSC	475

QY 42 -A- -LV- -P- -PA- -AL- -IS- 51
Db 476 EAVNDVSGSDSCSTEIVIKPPSPFIKLEPADIVRGTNALLQCEVSGTGPFPEISWFKDKQ 535
QY 52 -I- -I- -I- -I- -I- 52
Db 536 IRSSKKYRLFQSLVLEIFSPNSADVGEYECVAVNEVGKCGCMATHLLKBPPTFVKV 595
QY 53 -I- -Q- -KG- -L- -Q- 58
Db 596 DDLIALGQGVTLQAARVSEPISTVMKQGEVIREDKIKMSFNGVAVLIPDVQISF 655
QY 59 -Y- -VE-AE- -I- -I- -I- 63
Db 656 GGYTCLAENAGSQTSGELIVKEPAKIERAELIQVAGPATLEYTVAGTPELKPXW 715
QY 64 -V-S- -I- -N- -E-D- -G- -T-L- 72
Db 716 YKGRPLVASKKYRISFKNNVAQLKPYSAELHDSGOYTFEISNEVSGSCETTFTVLDRD 775
QY 73 -F- -D- -G-R- -P- -I- -E-SL 81
Db 776 IAPFFTPLRNVDSVVNGTCLDCKIAGSLPMRVSWFKDGEIAASDRYRIAFVEGTASL 835
QY 82 SLI- -DA- -V-MP- -DV- -V- -Q 93
Db 836 EIIRVMDNAGNFTCRATNSVGSKDSGALIVQEPSPFTKPGSKDVLPGSAVCLKSTFQ 895
QY 94 -T-R- -I- -I- -I- -I- -I- 95
Db 896 GSTPLTIRFKGNKELVSGSCYITKEALESLELYLVKTSDSGTCTCKVSNVAGVECS 955
QY 96 -Q- -Q- -A- -Y- -I- -I- 99
Db 956 ANLFVKEPATFVEKLEPFSOLLKGDATOLACKVTGTPPIKITWFANDREIKESSKRMSP 1015
QY 100 -R- -D- -D- -D- -K- -K- 102
Db 1016 VESTAVLRITDVGIEDSGEYMCBAQNEAGSDHCSSIVIVKESPYFTKPKPIEVLKEYDV 1075
QY 103 -LA- -Q- -Q- -Q- -Q- -Q- 107
Db 1076 MLLAEVAGTPPPEITWFKDNTILRSGRYKFTFQDHLVSLQLKFAADAGEYQCRVTNE 1135
QY 108 -A- -A- -A- -A- -A- 111
Db 1136 VGSSICSAVTLREPPSPFTKBIESTSLRGGTAAFOATLKGSLPITVWMLKDSDEITDD 1195
QY 112 -A- -A- -A- -A- -A- -AAS- 119
Db 1196 NIRMTFENNVAISLYLSGIEVVKHDKYVCOAKNDAGIQRCSALLSVKEPATITEEAVSIDV 1255
QY 120 -Q- -G- -SAK- -I- -I- -I- 125
Db 1256 TOGDPATLQVFGSGTKETAKWFKDQGLTGLSKYKISVTDVTSILKIISTEKKDSGEYT 1315
QY 126 -N- -G- -G- -G- -G- -G- 127
Db 1316 FEVONDVGRSSCKARINVLDLIIPSPFTKLKKWDSIKGSFIDLECIVAGSHPIQWFK 1375
QY 128 -E- -NTA- -N- -G- -E- -E- 134
Db 1376 DDOEISASEKYKPSFHDNTAFLEISQLEGTDSGTTCATNKAGHNQCSGHLTVKPPYP 1435
QY 135 -E- -N- -G-A- -H- -H- -H- 139
Db 1436 VEKPSQDVNPNTRVOLKALVGGTAPMTIKWFKDNKELHSGAARSVMKDDTSTSLFAA 1495
QY 140 -T-I- -A- -N- -N- -N- -N- 143
Db 1496 KATDSGTIYICQLSNDVGTATSKATLFVKEPPQPIKPPSPVLVRNGOSTTFBECQITGTPK 1555
QY 144 -N- -H- -T- -D- -D- -D- 147

Db 1556 IRVSWLDGNETAITQKHGISPIDGLATFOISGARVENSCTVVCBARNDAGTASCSELK 1615
QY 148 -MM- -EV- -D- -G- -D- -V- 155
Db 1616 VKEPPTFIRELKPVEVVKYSDVELECEVTGTPTTFFVTLKNNREIRSSKKYTLTDRVSFV 1675
QY 156 -E- -I- -PP- -N- -K- -K- 161
Db 1676 NUHITKCPDSTGEYQCIIVSNEGSCSTRVALKEPSPFIKIENTTVLKSSATFQST 1735
QY 162 -A- -V- -L- -R- -GH- -BS- 170
Db 1736 VAGSPISITWLKDDQILDENNVIISFVDSVATLQIRSVDNHSGRYTCYQAKNESGVER 1795
QY 171 -I- -C- -A- -W- -S- -D- 170
Db 1796 CYAFLLVOEPAQIVEKAKSDVTEKDPMTLECVVAGTPELVKWLKDGQIIVPSRYFMS 1855
QY 171 -E- -E- -V- -P- -F- -F- 173
Db 1856 FENNIVASFIOQVMKQDSGOYTFKVENDFGSSSCDAYLVLQNDONIPPSFTKLTWMDKVL 1915
QY 174 -I- -C- -A- -W- -S- -D- 177
Db 1916 GSSIHMECKVSGSLPISAQWFKDGKEISTSAKYRLVCHERSVSVLEVNLEDEDTANYTCK 1975
QY 178 -N- -P- -V- -S- -D- -D- 182
Db 1976 VSNVAGDDACSGILTVEKPPSFLVKPGRQQAIPDSTVEFKAILKGTTPPKIKWFKDDVEL 2035
QY 183 -LL- -AS- -GS- -GS- -GS- 188
Db 2036 VSGPKFCFLEGSTFSLNLYSDASKTGOYTCHVNDVGSDDCTTMLLVTEPPKFKVCKLE 2095
QY 189 -GDST- -A- -R- -IW- -N- -L- -S- -E- -N- -ST- 203
Db 2096 ASKIVKAGDSSRLCKEIKAGSPEIRVWFRNEHPSALDKYRMTFIDSVAVIQMNNLISTED 2155
QY 204 SG- -ST- -I- -I- -I- -I- -I- 210
Db 2156 SGDFICEAQNPAAGTSCSTKVIKPPVPFPPFIVETLKNABVLECSLGTTPPEVVM 2215
QY 211 -LR- -HC- -HC- -HC- -HC- 214
Db 2216 YKDKQLRSSKKYKIASKNFHTSIHLNVDTSDIGEYHCKAQNEVGSCTCVCTVKLKEPP 2275
QY 215 -IRE- -IRE- -IRE- -IRE- -IRE- 217
Db 2276 RFVSKLNSLTVVAGEPAELQASIEGAQPIFVQWLKEKEBIRESENIRITFVENVATLOF 2335
QY 218 -GG- -GG- -GG- -GG- -GG- 219
Db 2336 AXAEPANAGKYICQIKNDGGMENMATLMLVLEPAVIVEKAGPMTVTVGCTCTLECKVAGT 2395
QY 220 -QD- -QD- -QD- -QD- -QD- 221
Db 2396 PELSVENYKDKLLTSSQKHKESFYNNKISSILSVLVERQDAGTYTTFVQVNNVKSCTAV 2455
QY 222 -V-PS- -N- -N- -N- -N- -K- 226
Db 2456 VDVSRAVPPSFTRRLLKNTGGVLGASCILECKVAGSSPISVAFWHEKTKIVSGAKYQTTTF 2515
QY 227 -D-V-T- -SLD- -SLD- -SLD- -SLD- 232
Db 2516 SDNVCTQLNSLSDSDMGNYTCVAANVAGSDECRALVTVOEPPSPFVKEPEPELVLPKGV 2575
QY 233 -W- -W- -W- -W- -W- -S-E- 236
Db 2576 TITSVIRGTPPKVNNWFRGARELVKDCRCNIYFEDTVAELELFNIDISQSGEYTCVVSNN 2635
QY 237 -G- -T-LL- -A- -TG- -S- -Y- -DGF- 248

Db 2636 AGQASCTTRLFVKPEAFLKRLSDHSVEPGKSIILESTYTGTLPISTVTKOGFNITTS 2695
Qy 249 -----A-R----- 250
Db 2696 KCONIVTEKTCILEILNSTKRDAGQYSCBIENAGRDVCGALVSTLEPPYFVTELEPLEA 2755
Qy 251 -----I-W-----TK-----D-G----- 256
Db 2756 AVGDSVSLQCVAGTPEITVSWYKGTGKLRPTPEYRTYFTNNVATLVFNKNVINDSGEY 2815
Qy 257 -----N-L--AS-----TL-----G-----Q----- 264
Db 2816 CKAENSIGTASSKTVFRIQERQLPPSFARQLKDIEQTVGLPVLTCRLNGSAPIQVCWYR 2875
Qy 265 -----H-----K-----G----- 267
Db 2876 DGVLLRDHENLQTSFVDNVATLKILQDLSHSQYSCSASNPLGTASSSARLTAREPKS 2935
Qy 268 -----P-----I-----F-----A-LK-W-NK----- 276
Db 2936 PFDIKPVSIDVIAGESADFECHVTGAQPMRITWSKDNKEIRPGNYTITCVGNTPHLRI 2995
Qy 277 -----KG----- 278
Db 2996 LKVGKDSGOYTQATNDVKOMCSAQLSVKEPPKFVKLEASKVAKQESIQLECKISG 3055
Qy 279 -----N-FI-----L-----SA-----GV-D----- 287
Db 3056 SPEIKUSWFRNDELHESKYNMSPFINSVALLTINEASADSGDYICEAHNGVGDASCST 3115
Qy 288 -----K-----T-----T-----Iiw----- 293
Db 3116 ALTVKAPPVFTQKPSVPGALKSGDVILOCEISGTPPFVWVKDKQVRSKKFKITSKH 3175
Qy 294 -----D-----A-----H-----T-----G-----R----- 299
Db 3176 FDTNLHILNEASDVGEYHCKATNEVSGDTCSCSVKFKPPRFVVKLSDTSLIGDAVEL 3235
Qy 300 -----A-----K-----Q----- 303
Db 3236 RAIVEGQFQISVWVKDORGEVIRESENTRISFIDNIATLQLSPEASNSGKICQIKNDA 3295
Qy 304 -----F-----P----- 306
Db 3296 GMRECSAVLTLEPARIEKEPEMTVTGPNFALECVVTGTPELSAKWFKDGRELSADSK 3355
Qy 307 -----H-S-----A-----P-A-----L-----D-----V----- 315
Db 3356 HHITFINKVASLKIPCAEMSDKGLYSFEVKNSVGKSNCTVSVHSDRIVPPSFIRKLKDV 3415
Qy 316 -----W-----QS-----N----- 319
Db 3416 NAILGASVVECRVSGAPISVGFODGNEIVSGPKQCSFSFENVCTLNLSLLEPSDTGI 3475
Qy 320 -----N-----TP----- 322
Db 3476 YTCVAANVAGDECSAVLTVEPPSPETQSDSVEVLPCHSLFTSIVIRGTPPPKVKWFKG 3535
Qy 323 -----ASC----- 325
Db 3536 SRELVPESCNISLEDFVTELELFEVQPLESGDYSCLVTDNAGSACTTHLFVKEPATFV 3595
Qy 326 -----S-----T-----D-----M----- 329
Db 3596 KRLADPSVETGSPVLEATYTGTPPISVSWIKDEYLIOSERCISITMTEKSTILELEST 3655
Qy 330 -----C-----I-----HV-----CKL-G----- 337
Db 3656 IEDYAQSCLIEAGQDICEALVSLVEPPYFIEPLEHVAEIVGEPATLQCKVDGTGTPAIR 3715
Qy 338 -----Q-----D----- 339
Db 3716 ISWYKEHTKLRGAPYKMQFKNNVASLVINKVDHSDVGEYSCKADNSVGAVASSAVLVIK 3775

Qy 340 -----R-----P-----I-KT-----F-----Q-----G-----H----- 348
Db 3776 ARKLPFPFARKLKVHETLGGPVAFCRINGSPLQVSMYKDGVLKKDDANLQTSFVHN 3835
Qy 349 -----T-----N-----EV----- 352
Db 3836 ATLQILOTDQSHIGQYNCASNPLGTASSAKLILSEHEVPPFDLKPVSVDLALGSSGT 3895
Qy 353 -----NA-IK-W-D-----PTGN-----L-L-----A-SC-S----- 368
Db 3896 FKCHVTGTAPIKITWAKONREIRPGNYKMTLVENTATITVLKVGKGDAGQYTCYASNIA 3955
Qy 369 -----D-----DM-T-L-KI-----W-----SMK-Q----- 380
Db 3956 GKDSCSAQLGVQVQPPFIKKLPSRIVKODEFTRYECKIGSGPEIKVLWYKDET-ELQES 4014
Qy 381 -----D-----N-----C-----V-----H----- 385
Db 4015 SKFRMSFVDSVAVLEMHNLSDVEDSGDYTCENAHNAGSASSTSLKVKEPPIFRKXPHPIE 4074
Qy 386 -----D-----LO-----Q-----H----- 390
Db 4075 TLKGADVHLECELOQTPPHVSWMYKDKRELRSKGYKIMSENPLTSIHILNVDAADIGEY 4134
Qy 391 -----N-----KEI-Y-TI-----K 398
Db 4135 QCKATNDVGSDDTCVGSIALKAPPRFVKKLSLDISTVVGKEVOLQTTIEGAEPIVWVFKDK 4194
Qy 399 -----W-----S----- 400
Db 4195 GBIVRESNIMIWSYSENATLQFSRVEPANAGKYTCQIKNDAGMQECFATLSVLEPATIV 4254
Qy 401 -----P-----TG-----P-----G-----T-----N-----N----- 408
Db 4255 EKPESIKVTGTGDTCTLECTVAGTPELSTKWFKDGKELTSDNKYKISFFNKKVGLKINVA 4314
Qy 409 -----P-----N-----A-----N-L-----M-----L----- 415
Db 4315 PDSGVYSEVQNPVQKDSCTASLQVSDRTVPPSPTRKLEKETNGLSGSSVMECKYVGP 4374
Qy 416 -----ASAS-F-----DS----- 422
Db 4375 PISVSWFHEGNEISSGRKYQTTLTNTCALTVNMLESDSGDYTCIATNAGSDECSAPL 4434
Qy 423 -----TVR-----L-----W----- 427
Db 4435 TVREPPSFVQKPPMDVLTGNTVFTSIVKGTTPPFSVSWFKGSSSELVPGDRCNVLSDSV 4494
Qy 428 -----DVD-R-G-ICI-----H----- 436
Db 4495 ABLELFDVDTSGEYTCIVSNEAGKASCTTHLYIKAPAKFVKRLNDYSIEKPKLILEG 4554
Qy 437 -----TLT-----K-----H-Q-----B-P-----V-----YS-V----- 448
Db 4555 TFGTGPISVTWKNGINVTSPQRNITTEKSPILEIPSSSTVEDAGQYCNENASGKD 4614
Qy 449 -----A-----F-----SP-----D-----G-----R-----Y----- 456
Db 4615 SCSAQILILEPPYFVKQLEPVKVSGDSASLQCLAGTPEIGVSWYKGTCLRPPTTYKM 4674
Qy 457 -----L-----L-----A-S-G-----SF-----D-K-464
Db 4675 HFRNNVATLVFNQVINDSGEYICKAENSVEGVSASTFLTVOBQKLPSPFSRQLRDVQET 4734
Qy 465 -----C-V-----HI-----W-----NT-----QV-C 474
Db 4735 VGLPVVFDCAISGSEPIVSWYKDGKPLKDSPNVQTSFLDNTATLNI FKTDRSLAQYSC 4794
Qy 475 -----L-----H----- 476
Db 4795 TATNPIGSASSARLILTEGKNPPPPFDIRLAPDAVVGESADFECHVTGTQPIKVSWAKD 4854

QY 477 -----YL-N-----GQ-----V-----L----- 483
Db 4855 SREIRSGGKYQISYLENSAHLTVLKVDKSGGYTCYAVNEVGKOSCTAQLNIKERLIPP 4914
QY 484 -----L-----N-----L-GR-----S-----I-----C----- 491
Db 4915 SFTKRLSEVTEEGNSFKLEGRVAGSQPITVWYKNNIEIOPTSNCEITPKNNTLVQV 4974
QY 492 -----LYT-----L-----P-----H----- 498
Db 4975 RKAGMDAGLYTCKVNDAGSALCTSSIVIKPKPPVFDQHLTPVTSEGETVOLSCHV 5034
QY 499 ----- 498
Db 5035 QGSEPIRIOWLKAGREIKPSDRCSFASGTAVLELRDVAKADSGDYVCKASNVAGSDTT 5094
QY 499 -----L-----VV-----IP----- 503
Db 5095 KSKVTIKDKPAVAPATKKAADVGRGLFFVSEPOQSIRVVVEKTATTAIAKVGDDPIPNNVKT 5154
QY 504 -----L-----VA-----L----- 507
Db 5155 GKWRQLNOGRVFIHQKDEAKLEIRDTTKTDGLYRCVAFNEHGBIESNVNLQVDERKK 5214
QY 508 -----LE-L-----VLK 514
Db 5215 QEKIEGDLRAMLKKTPLK 5233

RESULT 2
T20992
hypothetical protein F15G9.4a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20992; T24733
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20992
A:Status: preliminary; translated from GE/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-5175 <W1>
A:Cross-references: UNIPROT:Q810L3; EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F1
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24733
A:Status: preliminary; translated from GE/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-5175 <W2>
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: CESP:F15G9.4a
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
1; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2981/3; 3033/1;
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/1

Query Match
Best Local Similarity 8.8%; Pred. No. 5.5e-53;
Matches 453; Conservative 49; Mismatches 8; Indels 4660; Gaps 389;
QY 1 M-----SI-----SS-----D-----E----- 7
Db 1 MGRSPSWLYGVGLLLATTCSVNDKNDPTGKSLAFVFDITGSMFDDLQVREGAAK 60
QY 8 -----V-----N-----F-----L-----VY----- 13
Db 61 IFKTVMAQREKLIYINWVPHDPIYLGEIINTDSTTFMRQLSKVYVHGDCPEKLTG 120
QY 14 -----R-----Y-----L-----QE-----SG-----F-----SHSAP-T 27

Db 121 ILKALQISLPSSFIYFTDARSKDVHLEDEVLTITQEKSSVVFVMTGDCGNRTHPGERT 180
QY 28 -----FG-I-----KS-----H-ISO-----S-NI----- 39
Db 181 YEKIAAASFGQVHLEKSDSVTVLEYYVRHAVKQKVHLMYFARERGGTVSRNIPVDKHL 240
QY 40 -----N-----GA-----LV-----P-P----- 46
Db 241 ELTISLGDKDDSDMLDIVLRDPEGRTVDKRLYSKEGGTIDLNKVKLIRLKPSPGVWTV 300
QY 47 -----A-----A-----LI-----S-I----- 52
Db 301 NTNSRLKHTIRVFGHGVDFKYGFASRPLDIRARPRVLNQDTYLLINMTGLIPPQTV 360
QY 53 -----I-----Q-----KGL-----Q-Y-----VE-----AE----- 63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYDEDDTEFMRIAT 420
QY 64 -----V-----S-I-----N-----ED-----GTLF----- 73
Db 421 AIGSVIVGGPAPFMSPIHQEFVGRDLNLSCTVESASAYTIYWKGTGEDIIGGLFVHNTD 480
QY 74 -----D-----G-----R-----P-I-----E----- 79
Db 481 TSVWTIPELSLKDAGEYECRVISNNGYSVKTRVETRESPPPEIFGVNRVSVPLGEAFLH 540
QY 80 -----S-----L-----SI-----I-----DA----- 86
Db 541 CSTRSAGEVEIRWTRYGATVFNPNTERPTNGTLKIHHTRADAGVVECMARNAGGMT 600
QY 87 -----VM-----P-DV-----V----- 92
Db 601 RMRDLIMEPPSVKVTPODVFYFNMREGVNLSCAEGDKPKPEVHWYFKGRHLLNDYKVG 660
QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----O----- 105
Db 661 QDSKFLYIRDATHDEGTVECRAMSGQA-RDITDLMLATPPKVEIIQNKMMVGRDVR 719
QY 106 -----Q-----Q-----Q-----A-----A-----A----- 111
Db 720 SPEKTIRKPKPKIRFWKNGKDLKPDYIKINEGOLHMGAKDEADAGAYSCVGENMAG 779
QY 112 -----A-----A-----A-----A-----A----- 114
Db 780 KDQVANLSVGRVPTTIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNP 839
QY 115 -----A-----A-----A-----ASQ----- 120
Db 840 RYTQLADGNLLITDAQIEDQGOFTCIARNTYQQSQSTTLMTVGLVSPVLGHVPPEQLI 899
QY 121 -----Q-----GS----- 123
Db 900 EGQDLTLSCVVVLGTPKPSIVMKDKKPVBEPTIKIEGSSLLRLRGNGPKDEKVTCTI 959
QY 124 -----A-----K-----N-----GE----- 128
Db 960 AVSPAGNSTLHINVLKPKPEFVKPEGGIVFKPTISGMDEKHVAVVNSTHVDLDGEFA 1019
QY 129 -----NT-----A-N-----G----- 133
Db 1020 IPCVSGSTPPPIITWYLDGRPTPNRSRDTFTADNTLIVRKADKSYGVVTCQATNSAGD 1079
QY 134 -----E-----E-----N----- 136
Db 1080 NEOKTIRIMTPMISFGSSFNWVDDLFTIPCDVYGPVKPVTWLLDDKPETEGVNE 1139
QY 137 -----G-----AH-----TI-A-N-----N----- 144
Db 1140 DGSLLTIPNVNEAHRGFTTCHAOAAGNDTRTTLTHTTPTINAENQEKIALQNDIVLE 1199
QY 145 -----HT-----D-----MME-----V-----D----- 152

Db 1200 CPAKALPPPVRLWTEYGEKIDSQLPHITIREDGALVLQNVKLENTGVFCVQSNLAGEDS 1259
QY 153 -----G-DV-----EIP-----P-----NK-AV----- 163
Db 1260 LSYLTIVHEKPKLISEVPGVVDVVGFTIIEPCRAATGVDEVITWNGIDLKMDEKPS 1319
QY 164 V-----LR-----G-----H-----E-----S----- 170
Db 1320 VDNLGLRIYEADKNDIGNVNCVVTNEAGTSQMTTHVDVQEPPIILPSTQNNNAVVGDR 1379
QY 171 -----E-----V-F-----I-C----- 175
Db 1380 VELKCYEASPPASVTWFRGIAIGTDTKGYVVESDGLTVIOASVEDATIYTCASNPA 1439
QY 176 -----A----- 176
Db 1440 GRAEALQVTVIASPDIKOPDVVTQESIKESHFSLYCPVFSNPLPQISWYLNKPLIDD 1499
QY 177 W-----N----- 178
Db 1500 KTSWKTDDKRLHVFKAKITDSGVYKCVARNAAGBSKSFQVEVIVPLNLDSEYKKKV 1559
QY 179 -----PVS----- 181
Db 1560 FAKGSEVTLGCPVSGFPVQINNVVDGTVVEPGKKYKGATLSNDGLTLHFDSSVSKQEG 1619
QY 182 -----D-L-LA----- 186
Db 1620 NYHCVAQSGNILDIDVELSVLAVPIVGEDDNLVFLGKDISLSCDLQTESDDKTFVWS 1679
QY 187 --GS-----G-----DS-----T-----A-R----- 194
Db 1680 INGSSEDRPDNVQIPSDGHRLEYITDAKPPENNGYKMCRTVNSAGKARTLLDLVLEPPV 1739
QY 195 -----IW-----N-----LS----- 199
Db 1740 EPVFEANOKLIGNNPILOQCVTGPKPTVIWKIDGNDVDKSWLPDESLSLLRIEKLTKG 1799
QY 200 -----EN-----S-----T-----S-GS-----T- 207
Db 1800 SAQISCTAENKAGTASRDFFIQNIAAPTFKNEGDQETIFRESEITILDCPVSUGDFOITW 1859
QY 208 --Q-----L-VL-----R-H-----C-----IR 216
Db 1860 MKQGLPLTENDALFTLDNTRLTILNARHEDIYTCVANNTAGQVSKDPDVVVQVLPKIK 1919
QY 217 -----EG-----G-----Q-DVP----- 223
Db 1920 NAVVTLEINEGERIILTCDAGNPTPTAKWDFNQGLDLPKEAVFVNNNHTVVVNNVTKYHT 1979
QY 224 -----SNK-----DV-----T-S-L-----D----- 232
Db 1980 GYVKVATNKVQAVTINHVTKPRFSGLTSELTVNLTRSIITLECDVDDAIGVGIS 2039
QY 233 W-N-----SE-G-TL-----L-----A-T-----GSY----- 245
Db 2040 WTVNGKPLAETDGVQTLAGRELHIVSAKTDDHGSYACTVNEAGVATKTNLFVQVPP 2099
QY 246 -----DG----- 248
Db 2100 TIVNEGGEVTVIENNSLVLPCVETGKPNPVVVTWKDGRPVGDLKSVQLSEGOQFKIVHA 2159
QY 249 -----A-----R----- 250
Db 2160 EIAHGSYICMAKNVGTABISFDVDIITKPMIQGKIKNIVTAIKGALPFPKCPIDDDKN 2219
QY 251 -----IW-----T-----K-D-G-N----- 257
Db 2220 FKQIILWLNYPIDLEADARITRLSNDRLTILNVTEDEQYSCRKNDAGENSFPDF 2279
QY 258 -----L-----A-S-TL-----GO-H----- 265
Db 2280 KATVLPPTIIMLDKDNKTAVEHSTVTLSCPATGKPEPDIITWFKDGEAIIHENIADIIP 2339
QY 266 ----- 265
Db 2340 NGELNGNQLKITRIKEGDAGKYTCCEADNAGSVEQDVNVNVIITPKIEKDIPSDYESQQ 2399
QY 266 -----K-G-P-----I----- 269
Db 2400 NERVVISCPVIARPPAKITWLKAGKPLQSDKFVKTSSANGQKLYFLKLRDSSKYTCIAT 2459
QY 270 -----F-----A-----L-----K-----W----- 274
Db 2460 NEAGTDKDRDFKUSMLVAPSFDEPNIVRRITVNSGNPSTLHCPAKGSPSPITITWLKGNAI 2519
QY 275 --N-----K----- 276
Db 2520 EPNDRYVFPDAGRQLQISKTGSDQGRYTCIATNSVGSDDLENTLEVIIPPVIDGERREA 2579
QY 277 -----K-G-N-----FI-----LS----- 283
Db 2580 VAVIEGSELSFCDNSNSTGVDEWQKGLTINQDTRLGDSFIQIPSSGKMSFLSARKSD 2639
QY 284 -----AG-----V-----DK-T-----TI-----I----- 292
Db 2640 SRYTCIVNPAGEARKLFDFAVNDPPSISDELSSANIQTIVPYYPVEINCVSQSGSPHPK 2699
QY 293 --W-----D-A-----H-----T-----G----- 298
Db 2700 VYWLFPDDKPLEPDSAAVELTNNGETLKIIVRSQVEHAGTYTCBAQNNVGRKARDPLVRVTA 2759
QY 299 -----E-----AK-----Q----- 302
Db 2760 PPHFEKEREEVVARVGDITMLLTCAABESSVPLSSVYVHHAHDESQNGVITSKYAANEKTLN 2819
QY 303 --Q-----F----- 304
Db 2820 VTNIQLDGFFYCTAVNEAGITKFFKILIVETPYFLDQQLYPIILKRLTLDSCATG 2879
QY 305 --P-----F----- 306
Db 2880 TPTPTILFMKDKRLNESDEVDIIGSTLVIDNPQKEVGRYTCIAENKAGRSEKDMVVEV 2939
QY 307 -----H-----S----- 308
Db 2940 LLPKLSKEWINVEVQAGDPLTLECIPTSGVHITWSRQFGDKGQDLMRAQSSDKSL 2999
QY 309 ---A-P-A-L-----D-----V-D----- 315
Db 3000 YIMQATPEDADSYSCIAVNDAGAEAVFQVNTVTPPKIFGDSFSTTEIVADTTLEIPCT 3059
QY 316 -----W-----Q----- 317
Db 3060 EGIPPEISFILDGKPILEMPGVYKQDLSLRIDNIKPNQEGRYTCVAENKAGRAEQDT 3119
QY 318 -----S-----N-----N-----T-----P----- 322
Db 3120 YVEISEPPRVWMASEVMRVVEGRQTTIRCEVFCNPEPVVWNLKDGEPYTSDDLQFSTKLS 3179
QY 323 -----A-----SC-----S-----TD-----M-C----- 330
Db 3180 YLHLRETTIADGGTYTCIATNKAGESQTTDDEVLVPPRIEDEBEEVLOGKEGNTYVWHCQ 3239
QY 331 -----I-----HV-----C-----K 335
Db 3240 VTGRPVVYTWKNGKEIEQFNPVLHIRNATRADEGKYSCIASNEAGTAVADFLIDVFTK 3299
QY 336 -----L-GO-----D-----RP-----I-K- 343
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHKPTISWLKGRPFNNMDNIIILSPRGDTLMILKA 3359
QY 344 -----T-----F-----Q----- 346
Db 3360 QRFDGGLYTCVATNSYSDSEQDFKNNVTKPYIDETIDQTPKAVAGGEIILKCPVLGNPT 3419

347 QY ---G---HT---N---E---VN---A-I- 355
3420 Db PTWTKRGDDAVPNSRHTIWNNDLKINSVTTEDAGQYSCIAVNEAGNLTHYAEVIG 3479
356 QY ---K--- ---M---D---P---T- 360
3480 Db KPTFVRKGNLYEVIENDTITMCGVTSRPLPSISWFRGDKPVLYDRYSISPDGSHITI 3539
361 QY ---G---N---L---LA---S---C-S 368
3540 Db NKAKLSDGGKYICRASNEAGTSIDLILKILVPPKIDKSNIGNPLAIVARTIYLECPSIS 3599
369 QY ---D---M--- --- --- --- 371
3600 Db GIPQPDVITKNGMDINWDSRVILQONNETFGIENVQVTDQRYTCTATNRGKASHDF 3659
372 QY ---TL---K---I---W---S---M--- 378
3660 Db SLDVLSPPEDIRHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLDITD 3719
379 QY ---K---Q--- ---D--- --- 381
3720 Db NVDISDDGRKLITISQASLENAGYTICIALNRAGEASLEFKVEILSPPEVIDISRNDVQPOV 3779
382 QY ---N---C-V-H---D---L---Q-Q---H--- 390
3780 Db AVNQPTIMRCVATGHPFSPSIKMLNGKEVTDDEINIRIVGQVQLILRTSDSHAGKWSVCV 3839
391 QY ---N---KE---I-Y-TT---KW--- 399
3840 Db AENDAGVKELEMDLVFTPPVSVKSDNPICALGETITLPCNASGNPYLPOLKWAAGSLI 3899
400 QY ---SP---T-G---P---GT---N--- 407
3900 Db FDSFPGARISLKGARLDIPLHKKTDVGYTCQALNAAGTSEASVSDVLVPPPEINRDGID 3959
408 QY ---N---P--- --- 409
3960 Db MSRLPAQQLSLTQLCQAQKFPQMRWTLNGTALTHTFGITVASTFTQINNVSLSDK 4019
410 QY ---N-A-NLM---LA---S--- 417
4020 Db GVYTCAENAVAGSNDLMYNDVQAPVISNGTKQVIEGELAVIECLVEGYPAPQWSLWR 4079
418 QY ---A-SPD--- --- 421
4080 Db NGRVETGVQVRYVTDGRMLTIEARSLDSGIYLCSATNEAGSAQOAYTLEVLVSPKII 4139
422 QY ---ST---VR-L---W---DV--- 429
4140 Db TSTPGVLTPSSGKFSILPCAVRGYPDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4199
430 QY ---D---R--- --- 431
4200 Db RHLIYEKTAKNADAGLTLEPPVQTIAPKISTSGNRYNGSEGTETVKEIBESSESFS 4259
432 QY ---G---I---C--- --- 434
4260 Db WSKNGVPLPSSNNLIFSEDYKLIKILSTLSDQGEYSCTAANKAGNATQKTNLNVGVAEK 4319
435 QY ---H---TL---TK--- 440
4320 Db IMERPRTQVVKGDQVTLNCEASGVQPAITWYKDNELLTNGVDETATTKKSVIFSSI 4379
441 QY --- --- --- 440
4380 Db SPSQAGVYTCAENWVASTBEDIDLIVMIPPEVPPVPMNVSTNPRQTVFLSCNATGIPEP 4439
441 QY ---H---Q--- --- 442
4440 Db VISWMRDSNIAIQNNEKYQLTGTLTALRNVLPPDDGFYHCIAKSDAGQKIATKLIWNKP 4499
443 QY ---E---P-V---Y--- 446

4500 Db SDRPAPIWECDEKGPKKTEYIMIDRGDTPDDNPQLLPWKDVEDSSLSNGSIAYRCMPGPR 4559
447 QY ---S-V---A-F---S---P---DG--- 454
4560 Db SSRTVLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPPHTITWAKDGLIEDSKFEIAY 4619
455 QY --- --- --- 454
4620 Db SHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAFINVDNNILPTPKPSSNQKNVAIICYER 4679
455 QY ---R---Y---LA---S---G--- 460
4680 Db NQAYSRLTWYNGVPMKPNLAGIHEFMNNGSLVILDTSSLKEGDELYTCKVNRHRSI 4739
461 QY ---S-F---DK---C-V---H-IW--- 469
4740 Db PHLTSAFEGVPEVKTIDKVEVNVNGDSVILDCVETSDPLTTHVVVWTKNDQKMLDDDAIYVL 4799
470 QY --- --- --- 469
4800 Db PNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQTLNVYGGSSRREAYKKENEDASTTTIT 4859
470 QY --- --- ---N---TQ--- 472
4860 Db TTSPTTTTETPLTTIIIPALITLPAKQYPTDDYHEGSANDDGGFGPTTQDSLFEFNPPLH 4919
473 QY ---V---C---L---H--- 476
4920 Db PEISVNTDCAGTINENGCDVKDKTHNLKILTGSHHCPEGFAMNPHTRICEDLDECAF 4979
477 QY ---LN---G-QVLN---LG---R---SI-C---L--- 492
4980 Db YOPCDECIYDGGFO--CNCPLGYELAEBCRDVNECESVRCECDKACFNQLGGYECID 5037
493 QY ---YTL--- ---P-H---H-L---V--- 500
5038 Db DPCPNYSLVDDRYCEPCENCTSTPIQVHMLAIPSGLPISHIATLTAYDKSRVLNDTT 5097
501 QY ---VI---PL---V---A---L---I---EL- 510
5098 Db YAISTDGAFGLARGMTSGPFTTICAKVRGHAQVMTNRVLAAAGDHHKVRVRAHSDHATNELH 5157
511 QY ---LVL 513
5158 Db APKTNFLVL 5167

RESULT 3

T43290

hemicentin precursor - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text_change 09-Jul-2004

C;Accession: T43290; T20993; T24734

R;Vogel, B.E.; Hedgecock, E.M.

submitted to the EMBL Data Library, June 1998

A;Description: Hemicentin is required for hemidesmosome mediated cell adhesion and germ-

A;Reference number: Z22396

A;Accession: T43290

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-5198 <VOG>

A;Cross-references: UNIPROT:O76518; EMBL:AF074901; PIDN:AAC26792.1

R;Sulston, J.

submitted to the EMBL Data Library, December 1994

A;Reference number: Z19355

A;Accession: T20993

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-5198 <WIL>

A;Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b

A;Experimental source: clone F15G9

R;Kershaw, J.

submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CBSP:F15G9.4b
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: him-4; F15G9.4b
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
A:Exons: 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/1

Query Match 70.9%; Score 2586; DB 2; Length 5198;
Best Local Similarity 9.9%; Pred. No. 6.2e-53;
Matches 423; Conservative 70; Mismatches 19; Indels 3770; Gaps 370;

QY 1 M-----SI-----SS-----D-----E----- 7
DB 1 MGRSPSLVGLGLLLATTCSSVNDKNDPTCKSSLAFFVDITGSMFDDLVQVREGAAK 60
QY 8 -----V-----N-----F-----L-----VY----- 13
DB 61 IFKTVMAOREKLIYINVPFHPDYLGEIINTDSTYFMRQLSKVYVHGGGDCPEKTLTG 120
QY 14 -----R-----Y-----L-----QE-----SG-----F-----SHSAP-T 27
DB 121 ILKALQISLPSSFIYFTDARSKDHYHLEVLNTIQEKSSVVVMTGCGNRTHPGFRT 180
QY 28 -----FG-----I-----KS-----H-----ISO-----S-----NI----- 39
DB 181 YEKIAAASFGVPHLEKSDVSTVLEVRHAKVKVHLMYAEARGGTVSRNIPVDXKLS 240
QY 40 -----N-----N-----GA-----LV-----P-----P----- 46
DB 241 ELTISLGDKSDNLDIVLRDEGRVTDKRLYSKEGGTIDLKNVKLRLKDPSPGVWTV 300
QY 47 -----A-----A-----A-----LI-----S-----I----- 52
DB 301 NTNSRLKHTIRVFGHGAVDKFGFASRLDRIELARPRVLNQDTYLLINMTGLIPPGTV 360
QY 53 -----I-----Q-----KGL-----Q-----Y-----VE-----AB----- 63
DB 361 GEIDLVDYHSHLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYEDENYEFMRIAFT 420
QY 64 -----V-----S-----I-----N-----ED-----GTLF----- 73
DB 421 AIGSVIVGGPRAFPSPHGFVGRDLNLSCTVESASAYTIYWKTCGEDIIGGLPFYHNTD 480
QY 74 -----D-----D-----G-----R-----P-----I-----E----- 79
DB 481 TSVWITPELSKDAGEYECRVISNNNGYVSKTRVETRESPEIFGVRNVSVPLGEAAFLH 540
QY 80 -----S-----L-----I-----DA----- 86
DB 541 CSTRSAGEVEIRWTRYGATVFNPNTERNPTNGTLKIHVTRADAGVYECMARNAGGMT 600
QY 87 -----VM-----P-----DV-----V----- 92
DB 601 RKMRLDIMEPPSVKVTFPQDVFYFMREGVNLSCAMGDPKPEVHWYPKGRHLLNDYKYQVG 660
QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q----- 105
DB 661 QDSKELYIRDATHDEGTGYECRAMSAGQA-RDITDMLATPPKVBIIQNMVMVGRGDRV 719
QY 106 -----Q-----Q-----Q-----A-----AAA-----AA----- 113
DB 720 SPECKTIRKPHKPKIRWFKNGKDLIKPDYIYKINEGQLHMGAKEDAGAYSCVGENMAG 779
QY 114 ---AAAA-----AS-----Q-----Q-----GS-----A-----KN----- 126
:::|

DB 780 KDVQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNNP 839
QY 127 -----GEN-----T-----A-----N----- 132
DB 840 RYTQLADG-NLLITDAQIEDQOQFTCIARNTYQQSQSTTLMVTGLVSPVGHVPPEQL 898
QY 133 -----G-----EE-----N-----G----- 137
DB 899 IEGQDLTSLCVVVLGTPKPSIVMIKDDKPVKEGPTIKIEGGSLRLRGGNPKDEGKYTC 958
QY 138 -----A-----H-----TI-----A-----N-----NH----- 145
DB 959 IAVSPAGNSTLHINVQLIKKPFVYKPEGGIKFTPTISGMDEKHVAVNASTHDLVDEGCF 1018
QY 146 -----T----- 146
DB 1019 AIPCVCVSGTPPPIITWYLDGRPIITNSRDTVTADNTLIVRKADKSYSGVYTCQATNSAG 1078
QY 147 D-----M-----M-----E-----VD-----GD-----V----- 155
DB 1079 DNEQKTTIRIMTTPMISPGQSSFNMMVDDLFTIPCDVYGDPKPVIITWLLDDDKPFTGCVN 1138
QY 156 E-----IP-----P-----N-----K-----A-----VVL 165
DB 1139 EDGSLTIPNVNEAHRGTFTCHAQNAAGNDRTVTLTHTTPTINAENQEKIALQNDIVL 1198
QY 166 -----R-----G----- 167
DB 1199 ECPAKALPPVRLWTYEGEKIDSLIPHTIRDEGALVQNVKLENTGVFCVQVSNLAGED 1258
QY 168 -----HE-----SEV-----F-----I-----C-----A-----WNP-----V----- 180
DB 1259 SLSYTLTVHEKPKILSEVPVGVVDVVGKFTIEIPCRATGVPEVIRTNKNGIDLKMDEKFF 1318
QY 181 S-D-L-L-L-AS-G-SG-D-D-S-TA----- 193
DB 1319 SVDNLGTLRIYEADKNDIGNYCVVTNEAGTSQMTTHVDVQBPPIILPSTQTNNTAVVGD 1378
QY 194 RI-----W-----NL-----S-ENST-----S- 204
DB 1379 RVELKCYEASPPASVTFWFRGIAIGTDTKGYVVEDGLTVIQSASVEDATIYTCASNP 1438
QY 205 -G-----S-----TO-----L-----V-----L----- 211
DB 1439 ACKAEANLQVTVIASPDIDKDPDVVTQESIKESHPSLYCFVFSNPLPQISWYLNDRPLID 1498
QY 212 -----R-----H-----CI-----R----- 216
DB 1499 DKTSWKTSDDKRLHVFRAKITDSGVYKCVARNAAGEGSKSFQVEIVPLNLDSEYKXK 1558
QY 217 -----EG-----G-----Q-----D-----V-----P-----SN----- 225
DB 1559 VFAKEGEEVTLGCPVSGPPVQINWVVDGTVVVEPGKKYKAGATLNSDGLTLHFDSVSVKQE 1618
QY 226 -----KDVTS-S-L-----D-----W----- 233
DB 1619 GNYHCAQSKGNLIDIVELSVLAVPIVGEDDNLEVLFLCKDISLSCDLQTSDDKTFVW 1678
QY 234 --N-SE-----G-----TL-----L----- 240
DB 1679 SINGSEDRPDNVQIPSGHRLYITDAKPENNNGKYMCRVTNSAGKAERTLTLDVLSPPVF 1738
QY 241 -----A-----TG-----SY-----D----- 246
DB 1739 VEPVFEANQKLGNNPIILQCVQVGNPKPTVIWKIDGNDVDSWLPDESLSLLRIEKLGT 1798
QY 247 -----G-----F-----A-----R-----I----- 251
DB 1799 KSAQISCTAENKAGTASRDPFFIONAAATFKNEGQETIFRESEITLDCPVSLGDFQIT 1858
QY 252 W-----T-----KO----- 255
DB 1859 WMKQGLPLTENDAIFTLDNTRLTILNANRDHEDIVTCVANNTAGQVSKDFVWVQVLPKI 1918
:::|

```
QY 256 -----GN-----L--A-----S--- 260
Db 1919 KNAVTLLEINEGEIIILTCDAENPTPTAKWDFNQGDLPKEAVFVNNHTVVVNNVTKYH 1978
QY 261 T-----L-----GO-----H--K----- 266
Db 1979 TGVYKCVATNKVGQAVKTVNHVTKRPFESGLTESELTVNLTRSTLECDVDAIGVGI 2038
QY 267 -----G-P-----I-----F----- 270
Db 2039 SWTVNGKPFLEAETDGVQTLAGRFLHIVSAKTDHGSYACTVNEAGVATKTNLFVQVP 2098
QY 271 -----AL-----K-----W----- 274
Db 2099 PTIVNEGGEVTVIENNSLVLPCEVTGKPNPVVTWKDGRPVGDLKSVQVLSEGOQPKIVH 2158
QY 275 -----N-----K-----KG----- 278
Db 2159 AETAHKGSYICMAKNDVGTAEISFDVDIIITRPMIQKIKNIVTAIKGALPFPKPIDDDK 2218
QY 279 NF-----I-----LS-----AG----- 285
Db 2219 NFKGQIWLRYOPIDLEADARITRLSNDRLTLNVLTENDRGQYSCRKNDAGENSFD 2278
QY 286 -----V-----D-KT-----TI-----I-W-D-A-H----- 296
Db 2279 FKATVLVPPTIIMLDKDKNTAVEHSTVTLSCPATKPKPEPDITWFKDGEAIIHENIADII 2338
QY 297 -----T-----G-BA-----K-----Q 302
Db 2339 PNGELNGNQLKIRIKEGDAGKYTCADNSAGSVEQDVNVNVTIPKIEKDGPSPYESQ 2398
QY 303 Q----- 303
Db 2399 QNERVWSCPVYARPPAKITWLKAGKPLQSDKEVKTSAHQKLYLFKLRDTSKYTCIA 2458
QY 304 -----F-----P-F-----H----- 307
Db 2459 TNEAGTKRDFKVSMLVAPDFDPNVRITVNSGNPSTLHCPAKGSPSTITWLKDGNA 2518
QY 308 -----S-----A-----P----- 310
Db 2519 IEPNDRVFPFDAGRLQISKTEGSDQGRYTCIATNSVSGDLENTLEVIIPPIDGERRE 2578
QY 311 --A-----L-D-----VD-WQ-----S-NT-----F-A----- 323
Db 2579 AVAVIEGFSSELPFCDNSGTGVVEWQKGLTINQDTRLGDSFTIQIPSSGKKMSFLSARKS 2638
QY 324 -S-----C-----S-TD-W-----CI-----H- 332
Db 2639 DSGRYTCIVRNPAEARKLFDFAVNDPPPSIDELSSANTQIVPYVPVEINCUVSGSPHP 2698
QY 333 -----V-----CK-----LG-Q-D----- 339
Db 2699 KVVYLFDDKPLEPDSAAELTNNGETLKI VRSQVEHAGTVTCEAQNNGKARKDFLVRVT 2758
QY 340 -----R-----P-----I-----KT- 344
Db 2759 APPHFEKEREVVARVGDWMLTFCNAESSVPLSVVYHHADESQVNGVITSKYAANEKTL 2818
QY 345 -----F-----Q----- 346
Db 2819 NVTNIQDDBGFYCYTAVNEAGITKFFKLVITETPFYLDQKLYPIILGKRLTDCSAT 2878
QY 347 G-----H-----T-N-EV-----N-A----- 354
Db 2879 GTPPTILFMKGDKRLNESDEVDIIGSTLVIDNPQKEVEGYTCIAENKAGREKDMVVE 2938
QY 355 I-----K-W-----DP-----T-----G-----N 362
Db 2939 VLLPPKLSKWINVEVQAGDPLTLECPIEDTSGVHITWRSQFGKQOLDMRAOSSDSK 2998
QY 363 LL-----A-SC-----S-----D-----DMTLKI--- 375
Db 2999 LYIMQATPEDADSYSCIAVNDAGBAEVQVTVNTPKIFGDSFSFSTEIVADTTLEIPCR 3058
QY 376 -----W-----SM-----KQ-----DN-----CV----- 384
Db 3059 TEGIPPPISWFLDKGPILKEMPGVTVYKQDLSLRIDNIRPNQGRYTCVAENKAGRAEQD 3118
QY 385 ----- 384
Db 3119 TYVEISEPRVVMASEVMRVEGRQTTIRCEVFGNPEPVVNMWLDGEPYTSLLQFSTKL 3178
QY 385 -----H-----D-L-----O-----Q-----H- 390
Db 3179 SYLHRETTLADGGTTCIATNKAQESQTTTDEVLVPPRIEDEERVLOQKEGNTYMHVC 3238
QY 391 -----N-KEI-----Y----- 395
Db 3239 QVTGRPVVYVWKRNGKEIEQFNPVLHIRNATRADEGKYSICIASNEAGTAVADFLIDVFT 3298
QY 396 -----T-----I-K-W-----SP----- 401
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMNDNIILSPRGDTLMILK 3358
QY 402 -----G-----T-----G-- 403
Db 3359 AQRFDGGLYTCVATNSYGSQDFKVVNVTKPYIDETIDQTPKAVAGGEIILKCPVLGNP 3418
QY 404 -P-----G-----T--NN----- 408
Db 3419 TPTVTWKRGDVAVPNDSRHTIVNNYDLKINSVTTEDAGOVSCIADVNEAGNLTHYAAEVI 3478
QY 409 --P-----N-----AN-L-M-----LASAS-F----- 420
Db 3479 GKPTFVRKGNLYEVIENDITITMDCGVTSRPLFSISWFRGDKPVYLYDRYSISPDGSHIT 3538
QY 421 -----D-----S-T-----V-R--L-- 426
Db 3539 INKAKLSGDKYICRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLECPI 3598
QY 427 -----W-----DV--D-R-----GI-----CI-----H- 436
Db 3599 SGIPQPDVITWTKNGBDINMTDSRVLAAONNETFGIENVQVTDQGRYTCATNRGKASHD 3658
QY 437 ---TLT--K--H-Q-----E-----P----- 444
Db 3659 FSLDVLSPPEFDIHGTQPTIKREGDTITLTCPKLAEDIADQVMDVSVWTKOSRALDGLT 3718
QY 445 --V-----YS-VA-----F-----SP-----D----- 453
Db 3719 DNVDISDDGKRLTISOASLENAGLYTCIALNRAGEASLEFKVEILLSPVVIDISRNDVQPO 3778
QY 454 -----G-----R-----Y-L-AS--G-S- 461
Db 3779 VAVNQPTIMRCVATGHPPFSIKMLKNGKEVTDENIRIVEQGOVLQILRTSDSHAGKWSC 3838
QY 462 -----F-----D--K-----C----- 465
Db 3839 VAENDAGVKELEWLDVFTFPVSVKSDNPIKALGETITLFCNASGNPYPQLKMAKGGSL 3898
QY 466 -----V-H-----I-----I 468
Db 3899 IFDSPDGAIRISLKGARLDIPLHKKTDVGYTCQALNAAGTSEASVSDVLVPEINRDMI 3958
QY 469 -----W-N-----T-Q----- 472
Db 3959 DMSRPLPAQOQSLTLOCLAQKFPVQMRWTNGTALTHTSTFGITVASDSTFIQINNVSLS 4018
QY 473 --V--C-----LHY-----L-NG--QV-----L- 484
Db 4019 KGYYTCYAENVAGSDNLMYNDVQVAPVISNGTKQVIEGELAVIECLVEGYPAPQVSWL 4078
QY 485 --N-----L--GR-----S-I-C-----LYTL----- 495
```

Db 4079 RGNRVETGVQGVRYVTDGRMLTIEARSLSIDSYLCSATNEAGSAQAYTLEVLVSPKI 4138
QY 496 -----P-----H-----497
Db 4139 ITSTPGVLTPSSGKFSKFCVAVRGYDDPIISWTLNGNDIKDGENGTIGADGTLHIEKAE 4198
QY 498 --HLV-----V--I--P-----LV-----AL---IE-----509
Db 4199 ERHLIYEKADNAGADTLFPVQTVAPKISTSGNRYINGSEGTETVIKCEIESSESSEF 4258
QY 510 -----L-----LV-----LK 514
Db 4259 SWSKNGVPLPSNNLIFSEDYK 4280

RESULT 4
T30192
probable peptide synthetase - Aureobasidium pullulans
C:Species: Aureobasidium pullulans
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
C:Accession: T30192
R:Peery, R.B.; Thorneswell, S.J.; Tobin, M.B.; Skatrud, P.L.
submitted to the EMBL Data Library, January 1997
A:Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aureoba
A:Reference number: Z20767
A:Accession: T30192
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-10797 <P>
A:Cross-references: UNIPROT:O94116; EMBL:U85909; NID:g4099310; PID:g4099313; PIDN:AAD005
C:Genetics:
A:Introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2
C:Keywords: carrier protein
F:1618-1688/Domain: acyl carrier protein homology <ACP1>
F:3682-3752/Domain: acyl carrier protein homology <ACP2>
F:5615-5685/Domain: acyl carrier protein homology <ACP3>
F:7503-7573/Domain: acyl carrier protein homology <ACP4>
F:9683-9752/Domain: acyl carrier protein homology <ACP5>

Query Match 70.8%; Score 2585; DB 2; Length 10797;
Best Local Similarity 8.8%; Pred. No. 3,1e-52;
Matches 457; Conservative 42; Mismatches 10; Indels 4680; Gaps 387;

QY 1 M-----SI--S-----S-----D-----6
Db 4302 MLNTQVIQFGSGSGKSGILYGMWGTEAAIHCTLPQGFVDLPAGTIGIPDITVSCFIV 4361
QY 7 -----E-V-----N-----FLV-----YR-----14
Db 4362 KPTESTKASOLELPIGELIGELVIGGHQADGYNLREBQTRAAF-VTHPKFGGLYRTGD 4420
QY 15 -----Y-----L-Q-----E-----SG-----P-----21
Db 4421 KARLHRNGTLECYGRISGQVKGQRVELGEIHAASKAGGCHAVIASVIGLLVLCI 4480
QY 22 --SH-----SA-----F-----T-F-----28
Db 4481 GDFHRVSSKDIKSAQKWLPAYMIPSDIVLDDFPYLPSPGKVDKRLTDYNSNTAQHVS 4540
QY 29 G-----IKS-----H-----I--SQ-----36
Db 4541 GSSDLSENAREITRIEIVSGVSDHSTLSAAGLDSLRAIQVASQLRQGCADIGALEL 4600
QY 37 -----SNIN-----G-----41
Db 4601 LSVSNVLDLVRKADESNINDSEKWKQTVHELRSVERDPESKAFVSGIEDVLP 4660
QY 42 -----A-LV-----P--P-----A-----AL-----IS--51
Db 4661 TPLQDAMLVETAKRPOAYCNEELRLTVSPKIPVERVRQALFALAQORHTALRSGFMPSGVSH 4720
QY 52 -----II-----Q--K--G-----L-QY-----VEAE--V-----64

Db 4721 CAYTVIMKTLVTSQFAHVKSFTTGWSTVNTRETLRLPLHFQYKCSGAEAILVAIHHALY 4780
QY 65 ---S---I-----NE-----D--G--70
Db 4781 DOWSVEVILEDLTLQNERTPERPSFGAVNKNFNLRRSEDOTSHLDTFWGEYLSDTVPGR 4840
QY 71 -----TL-----F-----D--74
Db 4841 LPNLSPKMPPOPLOSIQHTIEMDMETLQAAHSYSCSAHVFFQAAAYAILGYMGTEDT 4900
QY 75 -----GR--P--IES-----LS-L--I-----DA-----VM-----88
Db 4901 VFGTVFSGRTLPIVEIESVMGPLLSTLPTRINTLESRKFSVLSRLQEDNRKMRHSMTS 4960
QY 89 -----P-----D-V-V-Q-T-----R-----QQ--97
Db 4961 LADIKKACGFNPGEAVFDSIFVMQETARPDARAQTLNLNVEAHYLEFNLTLELEPTQOG 5020
QY 98 -----A-Y-----R-----DKL-----A--Q--105
Db 5021 VTKATYQSSLLPLQHVKTLLQQLDALVKIVVARPETHMNEISDQLPISVLSVANSEPOS 5080
QY 106 -----Q-----Q-----Q-----Q-----107
Db 5081 FVYKAGLSLVENHALNNSGGLALVFAHDIREGTSRMESLTYGELNTRANQLANYLISOG 5140
QY 108 A-----AA-----AA-----A-----113
Db 5141 AKRDELICVMEKSVSLYLILAAKAGCGYLPVPETPAARIRQILAEADVKFCITDSS 5200
QY 114 A-----A-----A-----A-----118
Db 5201 MAPVIADVSRCHIMNVDTDCSAQSGTGPQLDFKPTDIAVAVFTSTGTPKPGVLVTQEN 5260
QY 119 -----S--Q--Q-----G-----SA-----K--125
Db 5261 ILSNLEVLKSIYPPEGSRLQACNQAFDVSVEIFFTWTGWCCLCSASKDMFRDFEKA 5320
QY 126 N-----GE-----128
Db 5321 INELEITHLSLPTVAALTDPAPHRVKFLVTAGEAVTHHVHAGWAGKGLYQYGPSETT 5380
QY 129 N--T--A-----N-GE--EN-----GA-----138
Db 5381 NICTVNSAVESDHVINNIGPAPENTSAPVLTOGDDFQPLVPLGGLGELCFGGQVFRGYQN 5440
QY 139 -----H-----TI-----141
Db 5441 MPELTSKIINHPNYGRIVSRGDLGRLLPDGTILIQRTDDQKIRGQRIELGEISGCLL 5500
QY 142 ---AN-----N-----H-----145
Db 5501 QPPSVQNCIAEVIKTADKERLMAFIPSGYSKDSYSILOPKDNLEIIKSIYAHADNLP 5560
QY 146 -----T--DM-----M--E--150
Db 5561 AYMWPDALVPVSAIPQTSQGIKDKRLASDGSALTVEDLNAYSRGADDDTSELSATEQQ 5620
QY 151 -----VD-----G-D-V-----E-----I--P-----158
Db 5621 LASALADTLQMSQTSIGRSTSPFALGLDSVSAIRLATNLKKEYGVSIDVSYILKRPETIAR 5680
QY 159 --P-----NK-----AV-----VL-----165
Db 5681 LAPLLGSSKQTNPEFVTADCEAAVGSYLHDSVVSQLEHGQTVSQVLPCTPLQEAMLS 5740
QY 166 R--G-----H-----E-----S-E--V-FI-----174
Db 5741 ARDTSGSAYRNKTLFSLHGSVDKLCACWVEMQLQRHDLRTIFLSTEDSRFPVQAVLSQ 5800
QY 175 -----C-----A-----W-----177

Db 5801 WTLPMQECDDIPQLSTLLDSAKAGGDSIVDHSPPWKIQVYRSESTVYLLDMHLYDA 5860
QY 178 ---N---PVS---D 182
Db 5861 NAMSLLYEVEQLYKQDSAPVSKFPLNFMISTSVBEADALFRDQLREFVKKPRTD 5920
QY 183 ---L---LA---SG- 187
Db 5921 VKSGFGTITGRNLNYSKPMVETFLSKHSTTMLSITQAMWKTLLAASQSYSDVCCGVVSGR 5980
QY 188 --- 187
Db 5981 SVFVDGIESLVAPCFNTIPVRVDSKHSRNLGLVKALQVRNIDSLPYQLTPLRRIQAQAG 6040
QY 188 ---SG-D---ST--- 192
Db 6041 TNGKRLFDSLVLLQDPTDLDLSAIMELEGESGVMYMTSGTGLPKAVCLSHRAVTQSLL 6100
QY 193 A---R---I-W---N--- 197
Db 6101 AHDRTFSPSRFLQFASPTFEDVSVEIFFPWYRGATLVSVERNRLGLDLPGITSLNIDA 6160
QY 198 ---L---S---EN---S---T---SG--- 205
Db 6161 AELTPSVAASLVHRHENVPTLRALLTIGEMLNTQVIOQFGSGIGKGLYGMVGPTEAAI 6220
QY 206 ---ST--- 209
Db 6221 HCTLPQFGVDLPAGTIGIPDVTSCFIVKPTSTKSHASQLEILPITGEIGELVIGHOLA 6280
QY 210 ---V---L---R-H---C-I---R--- 216
Db 6281 DGVLNREEQTRAAFTVHPKFGGLYRTGDKARLHRNGTLECYGRISGGQVKLRQORVELGE 6340
QY 217 E---GG---SN---Q---D-V--- 222
Db 6341 IEHAASKAGGCHAVIASVIGLLVLCFICGPHRVSSKDIKSAQKWLPAWIPSDIVLLD 6400
QY 223 ---P---SN---KD---V---T--- 229
Db 6401 DFPYPSGKVDKRLFTDYNSTAHQVSGSSDLSENAREITRIIESVLGVSDHSTLSA 6460
QY 230 ---SL---D---W--- 233
Db 6461 AGDLSLRATQVASQLRQGCADLGALELLSVSNVLALDELVRAKADESNINDNSEKKWQ 6520
QY 234 --- 235
Db 6521 TVHELRSVVERDPESKAFVSGIBDVLPTCLQDAMLVETAKRPAQYCNELRLTVSPKIPV 6580
QY 236 E---G---TL---LA---TG--- 243
Db 6581 ERVRQALFALAQRHTALRGFMSGVSHCAYTQVIWKTLSQFAHVKSFTTGWSVTNRE 6640
QY 244 ---S---YD---G--- 247
Db 6641 TLLRPLHFQYKCSGAELVAITHALYDOWSVEVEILEDLETLLONERTPERSFGAVNK 6700
QY 248 -FA-R---I-W---TKD---G---NL--- 258
Db 6701 FFNLRSEDTSHLDFWGEYLS-DVTPGRPLNLSPKMPPQPLQSIQHTIEMDMETLRQA 6759
QY 259 ---A---S---TL---G--- 263
Db 6760 AHSYSCSAHVFOAAVAILLGFYMGTEDTVFGTSGRTLPIVEIESVMVGPLLSTLPTRI 6819
QY 264 ---Q---H---K-G-P---IFALKW--- 274
Db 6820 NTLESKFSVLSRLQEDNKKIMHSMTSLADIKKACGNFGEAVFDSIFV-WQETARP 6877
QY 275 ---N---K 276
Db 6878 DARAQTLNLLVEAHYLEFNLTLEPTQOGVKTKATYQSSLLPLOHVKTLLQOOLDALVK 6937

QY 277 ---K-G---N---F--- 280
Db 6938 IVVARPETHMNEISDQLPISVLNVANSEPOSFYKAGLSLVENHALNNSGGALVFAHD 6997
QY 281 ---I---LS---AG- 285
Db 6998 IREGTSMESLTYGBLNTTRANQLANYLISQAKRDELICVCMKESVSLYLSILAAVKAGC 7057
QY 286 ---VD--- 287
Db 7058 GYLPLVPETPAARIQILAEADVCKFLTDSSMAPVIADVSRCHIMNVOTDTCASQCTOP 7117
QY 288 ---K---T---I--- 291
Db 7118 QLDFTPTDIAYAVFTSGTTPKPKGLVLTQENTLSNLVLSKIYPVPEGSRLLOACNQAFD 7177
QY 292 ---I---W---D-AH--- 296
Db 7178 VSVFEIFFTWTYTMCLCSASKDWMFRDPEKAINLELEITHLSLTPVAALTDPAHIPRVKF 7237
QY 297 -T-GEA---K-Q--- 302
Db 7238 LVTAGEAVTHVHGAWAGKGLYQGYGPSETTNICTVNSAVESDHDVNNIGPAPENTSAPV 7297
QY 303 -Q---F-P---F---H---SA---P 310
Db 7298 LTQGDGFLVPLGGLGELCFGGQVFRGYQNMPELTESKIIHNPVNGRIYRSGDLGRLLP 7357
QY 311 ---AL---D---V---D---W--- 316
Db 7358 DGTILIQGRDQDKIRGORIELGBISGLLQFPVSQVONCAIEVIKTADKERLMAFWIPSG 7417
QY 317 ---Q---S---N--- 319
Db 7418 YSKDSYSILQDPKNLEETIKSIYAHADNLPAWMPDALVPVSAIPQTSQKIDKRRLAS 7477
QY 320 ---N---T-FA--- 323
Db 7478 DGSALTVEDLNAYSRGADDDTSELSATEQQLASALADTLQMSQTSIGRSTFFALGLDS 7537
QY 324 --- 323
Db 7538 VSAIRLATNLKREYGSIDVSOILKRPTIARLAPLLGGESSKQTNPEPVTADCEAAVGSYL 7597
QY 324 ---S-C---S---T---DM---C- 330
Db 7598 HDESVVSQLHEHGQTVSQVLPTQEQAMLARSATGSSAYRNKTLFSLHGSVDKLLKACW 7657
QY 331 -I---H---V---C---K---L---G--- 337
Db 7658 EYMLQKHDLRTIFLSTEDSRFPFQAVLSQWTLPMQECDDIPDQLSTLLDSAKAGDSI 7717
QY 338 ---Q---DR---PI-K--- 343
Db 7718 VDHSPPWKIQVYRSESTVYLLDMHLYDANAMSNLLYEVEQLYKQDSIAPVSKPFL 7777
QY 344 -T---P-Q---G---H--- 348
Db 7778 NFMISTSVBEADALFRDQLREFVKKPRTDVKSGFGTITGRNLNYSKPMVETFLSKHSTT 7837
QY 349 ---T---N---E---V--- 352
Db 7838 MLSLTQAMWKTLLAASQSYSDVCCGVVSGRSPVVDGIESLVAPCFNTIPVRVDSLSKHSR 7897
QY 353 N---A---I---K---W--- 357
Db 7898 NLGLVKALQRVNIDSLPYQLTPLRRIQAQAGTNGKLFDSLVLLOQDPTDLDLSAIVRLEG 7957
QY 358 ---D---PT---GNL---L---ASC---SD--- 369
Db 7958 ESGWMDFPCCIVELAPTNESYTLSLHFNRYLDDDEVVSNLHQACLSAFASCIIRYPSSDVSD 8017

QY 370 -----DM---TLK-----I-----WS-MK-Q----- 380
Db 8018 FIDFDADLVAGVLKPDTKHQPVAAKTNKRSKSGSGDESWSPLBLQIRAAYSAVSSAP 8077
QY 381 -----D-----N-----C----- 383
Db 8078 EDRIIRDITTYIKGLDISIAIQLANRLRKDGLLVQASDVMSPCSELASAVQSRQTPV 8137
QY 384 -----V-----H----- 385
Db 8138 LDERGDFGFKHYRGAALQSHRIATEKVASVRPCTPLQSGMLSEYTHSDGHQYFNHTF 8197
QY 386 -----D-----IQ-----OH----- 390
Db 8198 YAIADIDSKLQSAWSKVLEQHELLRTGFGVDDHEHPFVMLTYTEFDVIDLEIQASSR 8257
QY 391 -----N-----K-----E-----I----- 394
Db 8258 EGSVYSEKASESVKNNLHPFWRWSLLGVGRCLOQSAHAIIDAESLRLIMTDLQ 8317
QY 395 -----Y-----TI-----K-----WS-----P----- 401
Db 8318 SALSNGYVTRLTIDGALGHILSNSQADVESQRTFWSQKLSGAPVTRFPNMTFVRISDTE 8377
QY 402 -----LA-S-----A-S-----TG----- 403
Db 8378 AANVELVLNKRKLEARQELGVSQMSVQAAWARLLSAYTGESQVTFGVLSGRTPSA 8437
QY 404 -----P-GTN-----N-----P-NA-N----- 412
Db 8438 TADAAPCITTLFVSTNTAVDDSQFLKDLMSYNATIQKHQFTPLTIRNYAESTSEALPD 8497
QY 413 -L-----M-----LA-S-----A-S----- 419
Db 8498 SLFVYQRPMDVVDSSWKIIRKASVELAVSVEALSEDGLGLRLTTDPAQVPEHQK 8557
QY 420 -----F-D-----S-----T-----V-----R----- 425
Db 8558 IMLQMEVMIAGLLKFEADINTSVMSIIPKDPFIATDFKYLHEMTEASVKSYSDDRIAME 8617
QY 426 -----L-----W-----D-----VDRG----- 432
Db 8618 FVDAEDGQISSRHWTYRQLEDEANKIAHLIDRGVKPGDIIATSPDKCPEASFAFYGIL 8677
QY 433 -----C-I----- 433
Db 8678 KAGCAFCALDPTAPARKAFILEDNSARVLLTSDSIRSELRELTCQDIIIDLINFENKNEL 8737
QY 434 -----C-I----- 435
Db 8738 STSSVPVSLGAPSSVYLVTSQTTCTPKGCEITHDNAVLQVMSFKRLFKGRWTDSESWL 8797
QY 436 -----H-----T-----LTK----- 440
Db 8798 QFASYHFDVSVLEQFTWIVGMRLVCAPRDLILEDIAGFLDTMQITHLDTLTPSLGLLDP 8857
QY 441 -----H-----Q-----E-P----- 444
Db 8858 ALVPSLHKGVFTTGGESLKQDQINTWGDVGCLFNFGPTCTIGVTFPCVPKPKGKPSNI 8917
QY 445 -----V-Y-----S-V-----A-FSP-D-G-R 455
Db 8918 GWQFNVGCVVLAPGTQTPVLRGAIGELCISGLKVGKYNLRPELTADCF-PYLDAGFR 8976
QY 456 -Y-----LA-----SGS-F-----D-----K-----C-V-H----- 467
Db 8977 VYRTGDLVRLFDHSGSIDFLGRKDNQVKLRQRLDEIDEBAVIRKQDIIQDTVCIVAKHPK 9036
QY 468 -----IW-N-----T-----Q-----VC-----L-----H----- 476
Db 9037 QDKDQLIAFIAGINESKQKPELCPAESTRHLLIQTARAACEERLPGYVPTFLPQIRIP 9096
QY 477 -----Y-L-----N----- 479

Db 9097 LSVNNKVEKLRLQLYADLPFTTVIQTATQADSQQSLSDGEQKVAQALAEILLKIDNDLT 9156
QY 480 -----G-----QV--LL-N-----L-----GRS----- 489
Db 9157 PSSNIFSLGLSSISAIQPSKLLKASGFTTVQVATVVKNPFTISRLTKALATSTGRSGGETA 9216
QY 490 -----I-----C-----LY----- 493
Db 9217 DAKQVISACRQRHMGTVTRVLRCKADDIEAIAPCTPQQGIIISRSLSAESSLYFNSPKFN 9276
QY 494 -----TL----- 495
Db 9277 AQGVLDQLKEGAFNOALERTQILRTFFIETDDGYVQAVRKTGHLPMWLTLEVVDLASVDGV 9336
QY 496 -----P-----HH-----L----- 499
Db 9337 FAKRKQKWRSYNTSHLTVPFEIVIVRSGETFVSVDLHHALYDGNSTFIDLMNNVSKLYNS 9396
QY 500 -----V-----V-----I-----PL-----V-A-L-I 508
Db 9397 QEADFGKPFVDCIAGFGLRNVOGAKQFQWLDHLPDVKSASMPPLIDNPASHDVLCTASLDI 9456
QY 509 -----EL-----LV-----L-K 514
Db 9457 LNOADELRRLSLGTVQALVQATWVATLRK 9485

RESULT 5
I38344
titin, cardiac muscle [validated] - human
N:Alternate names: connectin
N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.1-)
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004
C:Accession: I38344; I38345; S20898; S20899; S63665; S37393
R:Labelit, S.; Gaute, M.; Paetore, A.
S:Science 270, 293-296, 1995
A:Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38344
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
A:Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:g1017424; PID:g1017425
R:Musco, G.; Tziatzios, C.; Schuck, P.; Paetore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix m
A:Reference number: I38345; MUID:95119041; PMID:7819249
A:Accession: I38345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:9602579; PIDN:CAA58243.1; PID:g602580
A:Note: conformation and properties are reported for a synthetic peptide corresponding t
R:Labelit, S.; Gaute, M.; Paetore, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380; PMID:1582406
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 13597-14200, 'I', 14202-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
A:Accession: S20899
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2;

A;Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
R;Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.
J. Mol. Biol. 256, 556-563, 1996
A;Title: Genomic organization of M line titin and its tissue-specific expression in two
A;Reference number: S63665; MUID:9617761; PMID:8604138
A;Accession: S63665
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 26729-26825 <KOL>
A;Cross-references: EMBL:X92412; NID:g1236761
R;Gautel, M.; Leonard, K.; Labeit, S.
EMBO J. 12, 3827-3834, 1993
A;Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentia
A;Reference number: S37393; MUID:94008990; PMID:8404852
A;Accession: S37393
A;Molecule type: mRNA
A;Residues: 26831-26926 <GAU>
R;Improtta, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A;Reference number: A66736; PDB:1IIT
A;Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R;Pfuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A;Reference number: A66201; PDB:1NCT
A;Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:l27867; OMIM:188840
A;Map position: 2q31-2q32
C;Function:
A;Description: structural protein forming filaments in striated muscle
C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro
C;Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein
F;24752-25008/Domain: protein kinase homology <KIN>
F;84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4068,
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
tatus predicted
F;16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18
21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248
F;26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 70.6%; Score 2575.8; DB 1; Length 26926;
Best Local Similarity 8.8%; Pred. No. 3.7e-51;
Matches 441; Conservative 57; Mismatches 11; Indels 4482; Gaps 379;
QY 1 M-----SISSD-----EV---N-----F-----L----- 11
DB 20024 MRENTETAEINLTKESVTADAGRYEITAAANSSTGTTKAFINIVLDRPPTGPWIS 20083
QY 12 -----V-----Y-----R---Y-L-QE-----S-----G--- 20
DB 20084 DITESVTLKWBPKYDGGSQVTNYLLKRETSATVWTEVSATVARTMMKWKLTGEEY 20143
QY 21 -----F-----P-----S-----H-----SA 25
DB 20144 QPRIKAENRGISDHIDSACVTVKLPYTPPGPSTPWTNVTRESITVGWHEPVSNGGSA 20203
QY 26 -----F-----T-----F-----GI-K-SH--- 33
DB 20204 VVGYHLEMKDRNSILWQANKLVIRTHFKVTTISAGLIYFPRVYAENAGVGKPSHPSE 20263
QY 34 -----IS---Q-S-----N---I-N-----G----- 33
DB 20264 PVLADACEPRNVRTIDISKNSVLSWQAPAFDGGSKITGYIVERDLDPGRWTKASET 20323
QY 34 -----IS---Q-S-----N---I-N-----G----- 41
DB 20324 NVTETQFTISGLTQSQYEFPRVARNAGVSNPSEVVGPICTIDSYGGVIDLPLEYTE 20383
QY 42 -----A-----L---V---P-P-----A-----LI----- 50
DB 20384 VKYRAGTSVKLRAGISGKAPTIEWYKODKEIQTNALVCVENTTDLASILKADRLNS 20443

QY 51 -----S-I-IQ--K-----G-----L-Q----- 58
DB 20444 GCYELKLRNAMASATIRVQILDKPGPGPIEFKVTVAEKITLLWRPPADGGAKITH 20503
QY 59 Y-VEA-E-----VS-----I-NE-----D----- 69
DB 20504 YIVEKRETSRVVSMVSEHLESCIITTTIKIKNNEIYFVRVAVNKYIGIEPLESDSVAK 20563
QY 70 -----G-----T-----L-F----- 73
DB 20564 NAFVTPGPGIPEVTKITKNSMTVMWSRDIADGSDISGVFLEKRDKSLGWPKVLKETI 20623
QY 74 -D-----G-----R-----PI-E-S-LSL-----ID-----A--- 86
DB 20624 RTRQKVTGLTENSQYRVCVNAAGOGPFSESEF-YKAAADPIDPPGPPAKIRIADST 20682
QY 87 --V---MP-D-----VW--Q-----T----- 94
DB 20683 KKSITLGNKSPYVDGSAVTVGVVEIROGEEBEWTVSTKGEVRTTEYVVSNLKPGVNY 20742
QY 95 -R-----Q-----QA----- 98
DB 20743 FRVSAVNCAGOGEPTEEMNEPVOAKDILEAPEIDLVALRTSVIAKAGEDVQVLIPFKGRP 20802
QY 99 -----YR-D-K-L--A-----Q----- 105
DB 20803 PPTVTRKDEKNGLSRDARYSIENTDSSLLTIPQVTRNDTGKYLTIENGVGEPKSTVS 20862
QY 106 -----Q-OA-----AA----- 110
DB 20863 VKVLDTPAACQKLVQKHSRGIVTLLDPPLIDGSGPIINYVIEKRDATKRTWSVVSHKC 20922
QY 111 -----A-----AA--A----- 114
DB 20923 SSTSKLIDLSEKTPFFRVLAEINEIGEPCEPTEPVKAAEVPAPIRDLNMDKSTKTSV 20982
QY 115 -----AA--A-----ASO--Q----- 121
DB 20983 ILSWTKPDGGSVITEYVVERKGEGTQWSHAGISKTCEIENVSQLKEQSVLEFRVFAKN 21042
QY 122 --G-S-----AK-----N----- 126
DB 21043 EKGLSDPTVITGVKLIITPEVDLSDIPGAQVTVRIGHNVHLELPYKGPSPISWLK 21102
QY 127 -G-----EN-T-----A-----N----- 132
DB 21103 DGLPLKSEBFVRFSKTENKITLSIKNAKKEHGGKVTVIDLNAVCRIVAPITVITLGPSPK 21162
QY 133 --G---E-----E-NG-----AHT-----TA----- 142
DB 21163 PKGPTRFDEIKADSVILSWDVPEDNGGGEITCYIEKRETSQTNMKNWCVSARTTFKVP 21222
QY 143 N-----N-----H-----T-D-M----- 148
DB 21223 NLVKDAEYQFVRAENRNVGSQPLVSSIIVAKHQFRIPOGPGKPYVINYVTSQMSLTWDA 21282
QY 149 -----ME----- 150
DB 21283 PVYDGGSEVTGPHVEKERNLSILWQKNTSPISGREYRATGLVEGLDYGQFVVAENSAGL 21342
QY 151 -----VD-G---E-I-T---PP-----NK- 161
DB 21343 SSFSDPSKFTLAVSPVDPGTPDYIDVTRETITLKNPPLRDGGSKIVGYSIEKQGNR 21402
QY 162 -----AV-----V----- 164
DB 21403 WTRCNFTDVSECOYTVTGLSPGDRVEFRIIARNAVGTISPPQSSSGIIMTRDENVPPIVE 21462
QY 165 -----LR-----G----- 167
DB 21463 FGPEYFDGLI IKSGESLRIKALVQGRPVPRVTFWKDGVIEIKRMNMEITNLVILGSTSLFVR 21522

QY 168 -H-----E-----S-----E-----V-----F-----I----- 174
Db 21523 DATRDHGVVTVAKNASAKAEIKVKVQDTGKVVGPPIRFTNITGCKMTLWMDAPLND 21582
QY 175 -C-----AW-----N----- 178
Db 21583 GCAPITHYIEKRETSRLAWALIEDCEAQSYYTAIKLINGEYQFRVSAVNKFGVGRPLD 21642
QY 179 --PV-----SD-----L----- 183
Db 21643 SDFVVAQIYVTPDAGCPEPSNITGNSITLTWARPESDGSSEIQOYILEREKKSSTRWV 21702
QY 184 -----L-----ASG-S-----G----- 189
Db 21703 KVISKRPISRETRFKVLTGEGNEYEFHVAENAGVGPASGISRLIKREPVNPPGPTV 21762
QY 190 ---D-S-T---AR-----I-----W-----N-----LS----- 199
Db 21763 VKVTDTSKTTVSLWSKVPDVGMEIIGYIEMCKTDLGDWHKVNABACVKTRYVTVDLQ 21822
QY 200 --E-----N-----S-----TSGS-----T-----Q-----L----- 209
Db 21823 AGEYKFRVSAINGACKGDSCEVT-GTIKAVDLRTAPELDIDANFKQTHVVRAGASIRLP 21881
QY 210 -----V-----LR-----H-----C----- 214
Db 21882 IAYQGRPTPTAVWSKPDNSLRADIIHTDTSFSTLTVENCNRNDACKYTLTVENNSGSKS 21941
QY 215 -----I-----R-----EGG----- 219
Db 21942 ITFTVKVLDTPGPGPIITFKDVTGRSATLMDAPLLDGGARIHHYVYVKEASRRSQWV 22001
QY 220 ---Q---D---VP---S---NK---:---:---DV--TS 230
Db 22002 SEKCTQIIFKVNDLAEGVPYFRVSAVNEYGVGEPEYEMPEPIVATEQAPPRRLDVVDS 22061
QY 231 -----L-----D-W----- 233
Db 22062 KSSAVLAWLKPDDHGSRTGYLLEMRQKSDLWVEAGHTKQLTFTVERLVEKTEYEFV 22121
QY 234 -N---SE-----G-T-LL--A---T-----G----- 243
Db 22122 KAKNDAGSEPREAFASVIIKEPQIEPTADLTGINTQLTCKAGSPFTIDVPISGRPAK 22181
QY 244 -SY-----D---G---F-----A----- 249
Db 22182 VTWKEEMRLKETDRVSIITTKDRTTLTVKDSMRGDSGRYFLTENTAGTKFTSVTVVVI 22241
QY 250 -R-----I-----W---TKDG-----NL-----ASTLG----- 263
Db 22242 GRPGVTGPIEVSSVSAESCVLSWGEPKDGGTETIYIIVEKRES--GTIAWLQVNSSVK 22299
QY 264 -Q---H---K-----G-----PI-----F-----A 271
Db 22300 RTQIKVTLTKYMEYSFRVSSNRFVSKPLESAPIIAEHFPVPPSPAPTRPEVYHVSANA 22359
QY 272 -----L---K---W---NK----- 276
Db 22360 MSIRWEPYHDGSKIIGYWEKKERTILWKENKVPCLCNYKVTGLVEGLEYOQRTY 22419
QY 277 ----- 276
Db 22420 ALNAGVSASEASRPIMAQNPDVADGRPEVTDVTRSTVSLIWSAPAYDGGKVKVGYIIE 22479
QY 277 -K-----G-----NF--I--LS-----AGV----- 286
Db 22480 RKPVSVEGDRWLKCNVTIVSDNFTVTALSEGDTVEFRVLAKNAGVISKSESTGPTV 22539
QY 287 -----DK----- 288
Db 22540 CRDEYAPPAELDARLHGLVTRAGSDLVDAAVGGKPEPKIITWKGKELDLCEKVL 22599
QY 289 --T---T-II----- 292

Db 22600 QYTGKATAVIKFCDRSDSGKYTLTVKNASGTKAVSMVKVLDSPGCGKLTIVSRVTQEK 22659
QY 293 ---W---D-A--H---T---GE----- 299
Db 22660 CTLAWSLPDGDGAETHYIVERRETSRLNWVIVEGECPTLSVYVTRLIKKNYIFRVRA 22719
QY 300 -----A-----K-----Q----- 302
Db 22720 VNKGPGVPESEPIVARNSTIPSPGPIEEVGTCKEHIIQWTKPESDGGNEISNYLV 22779
QY 303 -----QF-----P-----F----- 306
Db 22780 DKREKESLWTRVNKYVYVDYTRLKVTSLMEGCDYQFRVTAVNAAGNSPERSNFTSCR 22839
QY 307 -----HS--A--PALD---:---:---V-----D-W-Q--SN-- 319
Db 22840 BPSYTPGSPAPRVVDVTKHSISLAWTKPMYDGGTDIVGVLEMQBKOTDQWYRVHTNAT 22899
QY 320 --NT-----F----- 322
Db 22900 IRNTEFTVPDLKMGQKYSFRVAANVKNMSEYSESIAETIEPVERIEIPDLELADDLKTV 22959
QY 323 -----AS----- 324
Db 22960 TIRAGASLRMLVSVGRPPPVITWSKQIDLASRAIIDTTESYSLLIVDKWNRDAGKYT 23019
QY 325 -----C-----S-----T-D----- 328
Db 23020 IEAENQSGKSAVLVKVYDTPGCPSPVKVKEVSRDSVTITWEIPTDGGAPINNYIVEK 23079
QY 329 --M-----C---I---H----- 332
Db 23080 REAMRAFKYTVTKGSKTLYRISGLVEGTMHFRVLPENIYGIGEPCESTDVAVLSEVPL 23139
QY 333 -----V---CK----- 335
Db 23140 VPAKLEVVDTKSTVTLAWEKPLYDGSRLTGVYLEACKAGTERMKNVTLKPTVLEHTV 23199
QY 336 ---L-----G-----QD-R-P-I---KT----- 344
Db 23200 TSLNEGEQYLFRIRAQNEKGVSEPRETVTAVTVDLRLVPTIDLSTMPQKTIHPAGRPV 23259
QY 345 -----FOG-----HT-----N- 350
Db 23260 ELVIPIAGPPPAASWFFAGSKLRSESVTVETHKVAKLITRETIRTDTGETYLSLKNV 23319
QY 351 -----EVNA---IKW-----D--- 358
Db 23320 TGTSETIKVIILDKPGPTGPIKIDEIDATSIITISWEPPELDGGAPLSGVVVEQDAHR 23379
QY 359 --P-----T-GN-----LLAS---C-S----- 368
Db 23380 PGWLPVSESVTRSTFKFTRLTEGNEYVFRVAATNRFGIGSYLQSEVIECKSSIRIPGPE 23439
QY 369 ---D---D-MTL---K-----I-----W----- 376
Db 23440 TLQIFDVSRRDGMTLTWYPEDDGGSQVTGYIVERKEVADRWRVNVKVPVWTRYSSTGL 23499
QY 377 -----S---M---K-Q---DN----- 382
Db 23500 TEGLEYEHRVTAINARGSKPSRPSKPIVAMPDPIAPPKQPNPRVTDTRTTSVSLAWSVP 23559
QY 383 -----C-----V-----H----- 385
Db 23560 EDEGSKVTGYLIEQMKVDQHEWTKCNTTPTKIREYTLTHLPGQABYRFRVLACNAGPG 23619
QY 386 ---D-L---Q-----Q-----Q- 389
Db 23620 EPAEVPGTVKVTEMLEYDPYELDERYQEGIFVRQGGVIRLTIPIKGPFPPICKWTKEGOD 23679
QY 390 -----H---N--KE-IYTIK----- 398

```
Db 23680 ISKRAMIATSETHTELVIKEADRGDSGTVDLVLENKCGKAVY-IKRVIGSSENSDEGL 23738
Qy 399 -----WSP-----TG-----P-----CT-----N 407
Db 23739 EYDDIQRSVRVSRPDDGGADILGILERREVPAKAWYTIDSRVGTSLVVKLGKEN 23798
Qy 408 -----N-----P-N-----AN-----L----- 413
Db 23799 VEYHFRVSAENQGISKPLKSEPEVTPKPLNPPPPSPPEVLDVTKSVSLSWGRPKD 23858
Qy 414 -----M-----L-----A----- 416
Db 23859 DGSRVTGYVIERKETSTDKVRHKNKTQITMTYVTGLVPAEYQFRIIAQNDVGLSET 23918
Qy 417 S-AS-----F-----DS-T-----V-R----- 425
Db 23919 SPASEPVVCKDPKPSQPGELEILSKDSVTLQWEKPECDGKBEILGYWVEYRSGDS 23978
Qy 426 LW----- 427
Db 23979 ANKKSNERIKDKQFTIGGLEATEYEFVRPAENETGLSRPRTAMSIKTKLTSGEAPGI 24038
Qy 428 -----DV-----D-R-G-I----- 433
Db 24039 RKEMKDVTTKLGAAQLSCOIVGRPLDKWYRFGKELIQSRKYKMSDGRTHLTVMTE 24098
Qy 434 -----CT-----H-----TL----- 438
Db 24099 EQDEGVYTCIATNEVEVETSKLLQATPQHPGYPLKEKYGAVGSTLRHLHWYIGR 24158
Qy 439 -----SV-----A-P-----SP----- 452
Db 24159 PVPAMTFHCKQLQNSENITIENTHYTHVMKNVQRKTHAGKYKVLQSLNPGTVDAIL 24218
Qy 443 -----E-----P-----V-----Y----- 446
Db 24219 DVEIQDKPKPTGPVIEALLKNSAVISWKPADGGSWITNVVVEKCEKAEGAEQWLS 24278
Qy 447 -----SV-----A-P-----SP----- 452
Db 24279 SAISVTTTCRIVNLTENAGYFRVSAQNTFGISDPLEVSSVVIKSPFEPGAPKGTITA 24338
Qy 453 -----DG-----R-YL-----A-S-----G----- 460
Db 24339 VTKDCVWAKPPASDGGAKIRNYLLEKREKKQKNSIVTBEIRTFVSKNLIQLEY 24398
Qy 461 SF-----D-----K-----C----- 465
Db 24399 EPRVKCNLGESESEISEPITPKSDVPIQAPHFKEELRLNVRYSNATLVCKVTGHP 24458
Qy 466 ----- 465
Db 24459 KPIVKWYRQKEIIADGLKYRIQEPFGYHQHLLIASVTDDATVYQVRATNQGGSVSGTA 24518
Qy 466 -----VH-----I-W-----N----- 470
Db 24519 SLEVEVPAKTHLPKTLLEGMAVHALREVVSIKIPSGKPDPIVTQKQDLDLNNGHQY 24578
Qy 471 -----T-----Q----- 472
Db 24579 VIVTRSFVTLVPNGVERKDAGFYVVCACNRFGIDQKTVELDVADVPDPGRGVKVSASR 24638
Qy 473 -----V-C----- 474
Db 24639 DSVNLTWTEPASDGGSKIYIVIEKATTAERWLVRVQARETRYVINLFGKTSYQFRVI 24698
Qy 475 -----L-----H-----Y----- 477
Db 24699 AENKFGLSKSEPEPIITKEDKTRAMNYDEVDRETVSMTKASHSTKELYKYMIAE 24758
Qy 478 -L-----NG-OVL-----LN----- 485
Db 24759 DLGRGEFGIVHRCVETSSKTYMAKFVKVGTQDLVFKKEISILNTIARHNLHLHESPE 24818
```

```
Qy 486 ---L---G---R---S---I---C---L----- 492
Db 24819 SMEELVMIFEFISGLDIFERINTSAFELNEREIVSVHVQVCEALQFLHSHNIGHDPDIRP 24878
Qy 493 ---Y---T-----L-----P-H-H----- 498
Db 24879 NIYYQTRRSSTIKIIEFGQARQLKPGDNFRLLFTAPEVYVAPEVHQHDVVSTATDMWSLGT 24938
Qy 499 LV-V---I-P-L---V-----AL---IE---LVV----- 512
Db 24939 LVYVLLSGINPFLAETNOQIENIMNAEYTFDEEAFKEISIEAMDFVDRLLVKKRKRMT 24998
Qy 513 -----LK 514
Db 24999 ASEALQHPWLK 25009
```

RESULT 6

T42737

gp330 protein precursor - rat

N:Alternate names: megalin

C:Species: Rattus norvegicus (Norway rat)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004

C:Accession: T42737

R:Saito, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.

Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994

A:Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of t

A:Reference number: A58173; MUID:95024033; PMID:7937880

A:Accession: T42737

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-4660 <S>

A:Cross-references: UNIPROT:P98158; EMBL:L34049; NID:g561852; PID:g561853; PIDN:AAA51369.

A:Experimental source: strain Sprague-Dawley; kidney

C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-4660/Product: gp330 protein #status predicted <MAT>

Query Match

Best Local Similarity 70.6%; Score 2575.1; DB 2; Length 4660;

Matches 422; Conservative 62; Mismatches 27; Indels 3679; Gaps 367;

```
Qy 1 M-----S-----I-S-----SDEVN----- 9
Db 11 MLLLAIAACLEPVSQEGSGNFRCDNGYCI PASWRC DGT RCLD DTD TDEIGCP PRSCSG 70
Qy 10 -FL-----V-----Y-----YR----- 14
Db 71 LFLCPAEGTCIPSPSWVCDDEKDCSDGADRQNCAGTTC SAQQMTC SNGOCIPSEYRC DHV 130
Qy 15 -----Y-----Y-----L-Q--E- 18
Db 131 SDCPGSDERNCHYPTCDOLT CANGACYN T SORCQK VDCR DSSDEANCTT LCSQKE FEC 190
Qy 19 -SG-----F--SH-----S--AF-T--F--G--I----- 30
Db 191 GSGECILRAYCHDND CEDN SDERN CNVDT CGGHQFTCSNGCINQN WVCDDGDDCQDS 250
Qy 31 -----K--SH-----IS-----QSN----- 39
Db 251 GDEGCSNQSHRVCYPREWACPGSGRCISDKVCDGVDPCEP GDDENNVTSGRTCMGV 310
Qy 40 -----N-----GALV-PPAALI-----S-----I--I--OK----- 55
Db 311 CSVLNCEYCHQTPFGGECFCPPGHIINSNDRSTCIDFDQCIW GICDCKENRQGRHQC 370
Qy 56 -----G-L--Q----- 58
Db 371 LCEGYILRGQHKSSDSFSAASVFSNGRDL LVGLHGRNFRILAESKNRGMVGVDF 430
Qy 59 -Y-----V-----E-----AE--VS-----IN-----E----- 68
```

431 HYQKRVFTWDPQKVFSTDIINGLNTQBIILNSVDTPENLAVDWINNKLVLVETKVRNI 490
69 D-----G-----TL-----F-DG--R--- 76
491 DVVNLEGNQRBVITLITENLGHPRGIALDPTVGYLFFSDWGLSCQPKVERAFMDGSRKDL 550
77 -----P-----IE-----SL----- 83
551 VTTKVGWPAGITLDLVSKRVYVDSRYDIETVTDYDGIQKRTVARGGSLVPHFPGISLFE 610
84 --I--D--AVM-----PDVV-Q-----T-----RQ----- 96
611 EHVFFDTWKMAVMKASKFTETNPQVYHOSSLRPHGVTVYHALRQPNATNPGCSNNGGCA 670
97 QA-----YR-----D-----K--L--A-----Q--- 105
671 QVCVLSHRTDGLGRCKCEFGFELDDDEHRCVAVKNFLFSSKTAVRGIPPTLSTQED 730
106 -----Q-----Q-----A----- 108
731 VMVPVTGSPSFFVYDQHSVYFVYDLSKDIYKQKIDGTGKEVITANRLESVECLTF 790
109 -----A-----AA-----A--- 112
791 DMTSRNLWYTDGGLKSVTVLRLADKSRQIISLNNPRSIIVHPTAGYMFSLDWFPAKI 850
113 --A--A-----A-----A-----S-----Q--- 120
851 MRAWDSGLHMPVINTSLGWPNGLADWSASRLYWDAPFDKIEHSTLDGLDRKRLGHVD 910
121 Q-----G-SA-K-N-----GE 128
911 QMTHPFGLTVFXDNVFTDWRLGAIIRVRKSGDGMTVIRRGISSVMHVKAYDADIQTGS 970
129 N-----T-ANG-----E----- 135
971 NYCSTTHANGDCHGFCFPVPNQRVCGPYGMKLQDQMTCEGDPAREPPTQCGSLSF 1030
136 --NG-----AH-----T-----I-----AN 143
1031 PCWNGKCVSPFRCDGVDDCHDNSDEHQCVFNNTCSPAFACVRGQCIPGWCHCDRON 1090
144 -----N--H-T-----D-M-----ME-V-----DG-DVE----- 156
1091 DCLDGSDEQNCPTHATSTCPSFTCDNHVCIPKDWCDTNDSCDSD-EKNCOASGT 1149
157 -----I--P-----P-----N-----K-A-----V----- 163
1150 COPTQPRCPDHRICISPLYVCDGDKDADGSGDEAGCVLNTCTSAQFKCADGSSCINSRYCD 1209
164 -VL-----R-G-H-----E-----S-E-----V- 172
1210 GUYDCRDNDEACPRPQMGCHPDBFCQCGDGTCTPNTWECGHPDCIHGSDHTGCVP 1269
173 -----FIC-----AW-----N-----P----- 179
1270 KTCSPHFLCDNGNCYKAWICDNDRCRMSDEKDCPTOPHCPSTQWQCPGYSCTINL 1329
180 -----V-----S-----D-----L-----LA----- 185
1330 SALCDGVFCPNGTDESPLCNQDSCSHFNGGCTHQCWQGPFGATCLCPLGYQLANDTKC 1389
186 -----S-----GS-----G----- 189
1390 EDINECDIPGFCQHCVNMRGSPRCADPEYTLSDGRTCKVTGSENPLLWASRDKIIV 1449
190 -----DS-TARI-----W-----N-----L 198
1450 DNTAHTNHLVSLVQDVSFVALDPSVTCRVFWSDLLQKTSWFQNGTDKRVVHDSGL 1509
199 S-E-----N--S-----T-----S-----GS-----T--Q-LVL--R----- 212
1510 SVTEMIADWIGRNLWYVALETIEVSKIDGSHRTVLISKVTKPRGLADPRMGDNVM 1569

QY 213 -----H-----C-----I----- 215
1570 FMSDWGHHFRIERASMDGTMTRTVIVQEKIYWPCLSIDYPNRLIYFMDAYLDVIEBCDYD 1629
216 -----R-----E-----GG----- 219
1630 GHNRRQVIASDLVLRHHPHALTLFEDFVYWTDRGTROVMQANKWHGQNSVVMYSHVQPLG 1689
220 -----Q-----D-----V-----P--- 223
1690 ITAIHPSROPPSRNPNPCASCSHLCLLSAQAPRHYSACPSGNLSDDSVNCVRGQDPFL 1749
224 -S--N-----K-D-V-TS-----LD-----W--N-SE-----G 237
1750 MSVRDNIIFGISLDPEVKSNDAMVPISGIGHGVDFEFDDEQFIYVVENPGEIHRVKTGD 1809
238 -----T-----L-LA-----T-----G-S-Y-----D----- 246
1810 SNRTVPAPLSLGLSSGLALDWMVSRNIYVTPASRSIEVLTLLKGDTRYGKTLIANDGTPL 1869
247 --GF-----AR--I-W-----T-----K-----D-----GNL----- 258
1870 GVGFPVGIADVARGKLYWSDHGTSGVPAKIASANMDGTSKLILFTGNLQHLLEVVTLDI 1929
259 -----A--S-----T-----LG-----Q----- 264
1930 QEOKLYWATSRGVIERGNDGTTERMLLVHHLAHPGLVYGSFLYYSDEQYIEVERVDK 1989
265 -----HK-----G-----P-----I----- 269
1990 SSGNNKVLDRNDVPYLRGLRVYHRRNAADSSNGCSNPNACQICLPVPGMFSCACASG 2049
270 -----P-----A-----LK----- 273
2050 FKLSPDGRSCSPYNSFMVVMPLPAVRGFSLELSDHSEAMVPVAGQGRNVLHADVDVANGF 2109
274 --W-----N-----KK-G-NF-----I-----L--SAG----- 285
2110 IYWCDFSSVSRSGIRRIKPDGNSFTNVVYIGANGIRGVALDWAAGNLYFTNAPVE 2169
286 -----VD-----K-----T--T-- 290
2170 TLIEVLINTYVRRVLLKVSVDMPRHIIVDPKHYLFWADYGQKPKIERSFLDCTNRTVL 2229
291 -----I-----I-W-----D--A--H-----T--GE 299
2230 VSEGIPTPRGLAMHDHTGYIYVWDDSLDIARIHLDDGESQVVRYSRYPTPYGITVFE 2289
300 -----A-KQ-----Q----- 303
2290 SIWVDRNLKKVQASKQPGNTDPPVIRDKINLLRDVTIFDEHAQPLSPAELNNPCLQ 2349
304 -----F-----P-----F-----H 307
2350 SNGGSHFCFALPELTPRCGCAFGTLGNDGKSCATSOEDFLIYSLNLSRLHFPDRDH 2409
308 SAP-----ALD-----V- 314
2410 SLFPQVISVAGTAIALDYDRNRNRIFFTOQLNSLRQISVLSYSGSSPTVLLSNIGVT 2469
315 -----DW-----Q-----SN----- 319
2470 DGIADFWINRRIYSDFSNQITNSMAEDGSRAVIRVSKPRAIPLDPCRGYMYTDMGT 2529
320 -----NT----- 321
2530 NAKIERATLGGNFRPFIIVNTSLVWPNGLALDLETLLYWADASLQKIERSTLTGTNRVV 2589
322 -----F-----A-----S----- 324
2590 VSTAFHSFGLTVYGOYIYWTDLYTRKIYRANKYDGSDLVAMTTRLTLPQSGISVTVKTOR 2649

QY 325 --CST--D--M--C--I--H--V--CK--L 336
Db 2650 QCCSNPCDQFNGGSHICAPGNAECQCPEGWYLANDNKYCVDTGTRCNQLQFTCL 2709
QY 337 -G--QD-- 339
Db 2710 NGHCINQDKNDNDGSDGDELPTVCAPHTCRSTAFTCGNGRCVPHYRYCDYNDGCD 2769
QY 340 --R--P--I--K--T--F-- 345
Db 2770 NSDEAGCLFRNCNSTTFTCSNGRCIPLSVYVCGINNCHNDTSDERKNCPPPTCPDFTK 2829
QY 346 -Q--G--HT--NEVNA--IK--W-- 357
Db 2830 CQTNICVPRAFLCDGNDGSDENPIYCASTCRSNEF-QCLSPQRCIPSYWFCDE 2888
QY 358 --D--P--T--G--NLL--AS--C--S--D--DM-- 371
Db 2889 ADCADGSDPTCGHSVNTCRASQFQCDNGRCISGNWVCDGNDGMSDEDDQRHHCCLQ 2948
QY 372 --T--L--K--I--W--S--M-- 378
Db 2949 NCSSTQTCVNSRPPNRRRCIPQYVWVCDGADCDALDELQNMTCRCSAGEFSCANGRCV 3008
QY 379 KQ--D--N--C--V--H--DLO-- 388
Db 3009 RQSPCRDRNDGSDYSDERGSPPCCHANOFTCONGRICPRFPVFCDENDGSDGDEQEH 3068
QY 389 --Q--HN-- 391
Db 3069 LCHTPEPTCLPQRCNDGHCIEGRVNCNVHVDGSDNSDEKGCINECLDSSISRCDHNC 3128
QY 392 --KE-- 393
Db 3129 TDTITSFYCGLPGYKLMDSKRSQVDIDECKSPQLCSQKCNVWGSYICKCAPGYREP 3188
QY 394 --I--Y-- 395
Db 3189 DGKSCRQNSIEPYLIFSRYIRNLTTDSSYSLLIQLGNGVVALDFDRVEXKLWIDA 3248
QY 396 --TI--K--W--S-- 400
Db 3249 EKQIIRMFNLKNTRETIINHLRRAESLAVDWVSRKLYLWLDAILDCLFVSDLEGRHKM 3308
QY 401 --PTG--P--CTN--N-- 408
Db 3309 IAQHCVDANNTFCFEPHPRGIVLHPQRGHVYVADGWVHAYIGRIGMDGTNKSVIISTKIEW 3368
QY 409 PNA--N--LML--ASA--SF--D--S--TVRLW-- 427
Db 3369 PNAITIDYNDLLYWADAHLGYIEFSDLEGHRRHTVVDGSLPHFPALTIPEDTV-FWTDW 3427
QY 428 --D--V--D--R--G--I-- 433
Db 3428 NTRTVEKNGYDGSRVLVNTTKFPDIHVYHPYQPIMSNPCTGNGGCSHLCLIKAG 3487
QY 434 --CI-- 435
Db 3488 GRGFTACPDPTQVQLRDLTCLWPMCSSTQFLCGNNEKCIPIWKCQKQCSGSDGDEP 3547
QY 436 --H--T--L--T--KH-- 441
Db 3548 DLCPHRCFLGQFCRDGNCSTSPQALCNARQDCADGSDDEDRVLCHEHRCSNEWQCANR 3607
QY 442 --Q--E--P--V--YS-- 447
Db 3608 CIPQSWQCSVNDCLNDSDEDTSHCASRTCPGQFKCNGNRCIPQSWKCDVDNDGCDYSD 3667
QY 448 --V--A--FS--P--D--G-- 454
Db 3668 EPIDECTAAYCNDNHTFESCKTNYRCIPQWAVCNGFDDCRDSDSGGCVSPCHPSGDF 3727
QY 455 --RY--L--A--S--G 460

Db 3728 RCANHHCIPLRWKCDGTDGCDNSDENVCVPRCSSEFRCADQOCIPSRWVCDQNDG 3787
QY 461 --S--F--DK--C-- 465
Db 3788 DNSDERDCWKTCHEPHFOCTSGHCVKALACDGRADCLDASDESACPTFRFPNGTYCPAA 3847
QY 466 -V--H--I--W--NT--Q-- 472
Db 3848 MFECKNHVICQSWICGENDCVDSDEBIHLFCNIPCESPQFRCDNSRCVYGHQLCNG 3907
QY 473 V--C--LH--Y--L--NG--Q--V--L--LNLG 487
Db 3908 VDDCGSDGDEKEHCKRPHKPCCTDEYKCSNGNCISQHYVCDVNDVNDGDLSDETGCLG 3967
QY 488 --RS--TC--L-- 492
Db 3968 DNRCTAENICEQNCCTQLSSGGFICSCRPFGKPSDKNSQDINECEEFGICPQSCRNSK 4027
QY 493 --Y--T--LP--H-- 497
Db 4028 GSYECFCVDFGFKSMSTHYGERCAADGSPPLLLPENVRIRKYNTSEKSEYLEEHEHQ 4087
QY 498 --H--L--VV--I--PL--V-- 505
Db 4088 TIDYDWDPEHIGLSVYVTVLAQGSQFGAIKRAYIPNFESSGNNPIREVDLGLKYLQMPD 4147
QY 506 --A--L--IE--L--LV--L--K 514
Db 4148 GLAVDWVGRHYWSDAKSORIEATLDGRYKWLITTDQPAIAVNEPK 4197

RESULT 7

T15789

hypothetical protein C41A3.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

R;Accession: T15789

R;Bentley, D.

submitted to the EMBL Data Library, November 1995

A;Description: The sequence of C. elegans cosmid C41A3.

A;Reference number: Z18404

A;Accession: T15789

A;Status: preliminary; translated from GB/EMBL/DBBJ

A;Molecule type: DNA

A;Residues: 1-7829 <BEN>

A;Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA831

C;Genetics:

A;Gene: C41A3.1

A;Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/2;

/2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3

C;Keywords: carrier protein; phosphopantetheine; phosphoprotein

F;6547-6616/Domain: acyl carrier protein homology <ACP>

F;2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 70.5%; Score 2574.2; DB 2; Length 7829;

Best Local Similarity 7.9%; Pred.No. 3e-52; Indels 5238; Gaps 381;

Matches 457; Conservative 46; Mismatches 10; Indels 5238; Gaps 381;

QY 1 MS--ISS--D--E--V--N--F--LV-- 12

Db 1 MSACLVGWSWAKTPNGEDGHEMAENIFLNRNNTALLPVEKNYLSDFREKHPEKVALVDGI 60

QY 13 --Y--Y--R--Y-- 15

Db 61 EYFDDQYFGTGESEAIACMDPQOQMLMQGVIKGLGNAGITLEMASEARVAVYTAAMCYDK 120

QY 16 --L--Q--E--SG-- 20

Db 121 DLLPDDQWATGNSASVMCGRITYFLNSRGAAGVETACSSSLVAFHLARQAIQSGETKL 180

QY 21 --F--SH--SA--F--TF-- 28

Db 181 ALVCGANHVGRSFGHSLYNSHMVSPNGRLAAFDRSANGFVRAESFAVAVLCSKQFAEENN 240
QY 29 -----G-----I-K-S-----H----- 33
Db 241 LLHCECVGSFNSDGTPLSTAPNPISQYEVQLEALKNIDKDSVQLVTCHGTGTLKGDQ 300
QY 34 -----I-----S-----OS-----N----- 38
Db 301 VELTAINRSPKSDIRVMSPKSGMHGEGAAGLIGVLQSLYSMHQGIIPNQLHLELPSED 360
QY 39 -----IN-----G-----ALV--P--P-----A-----A-- 48
Db 361 GEDKMGFVNEEMELNRVAISSYFGGTNACATIEKPEKPSLVQKESYAESNVFLSAKS 420
QY 49 -----L-I-----S-----I-----I-----QKQ-- 56
Db 421 HESLKLQIEEYTOFMAQDSAMEDILYTVNERKTKYDFRAAVPGKDNEELARKLQGDYS 480
QY 57 -----LQ-----Y-----V-----EA-----E----- 63
Db 481 LTNLQESTFEVERGEKNEKLLMLRLMYEKNETFHTVDKYCKLAETCGPFEARTALFFPF 540
QY 64 -----VS-I-----N----- 67
Db 541 KLTLELTYNVSRLISSMATFELLVQNTLPNKLKGLQIFCLAVAKVITPESAVOLI 600
QY 68 -----E-----D-----G-----T----- 71
Db 601 KGVVAEANLTDLGDIELKSSKIPIBIOHLKSTKKILPIHISGELKETAKPNLMTPIVN 660
QY 72 -----L-----FD-----GR-----P-----I 78
Db 661 GBEILLEDPVRKVOKLICQLFACGFPDPAVKFRGRIVKTPTYSFLKQFPWQVQTAMTNI 720
QY 79 -----E-----SLS-----L-----I-----D-----AV----- 87
Db 721 VDEQTNSSLSABIESTVARTIVKQFLDIBEDDINLLETGAVDLSIEMVEAFGTAVNQT 780
QY 88 MP-----T-----R-----Q-----D--VV--Q----- 93
Db 781 MPFDLEAYPITLINIVDFLKLTVTPVTKATTSIHKKTSELSTSDINVIACDYQPAGVE 840
QY 94 -----T-----R-----Q----- 96
Db 841 GEKELWDLTSLRTLTGKSIDIRKKQCEGDAGLEVGLLKQDISMPDNSFFAIKQDEAFL 900
QY 97 -----Q-----AY----- 99
Db 901 DPQHRLLNAAYNALKEKSLTSPDADFLAISAHSEYRALAEKHINELDERLMMGTVHS 960
QY 100 -----R--D-----KLA-----Q-Q----- 106
Db 961 MVAGRLAVLMGIRGRAMIVDTTCSSVATALEMAVKSIREGRPAIVATSQLIOSKWLVS 1020
QY 107 -----Q-----Q-----A-----A-----A----- 111
Db 1021 LKTLDDHSTNSFSDGSGFCRSDGVGVIILKTAEGDSAVIKISAKSHHCCAVMTVPV 1080
QY 112 -----A-----A--A--A--A-----A-----AS----- 119
Db 1081 SSISQLLEBAGFSYVEGHGTATSGDSDAESMAYQKGLSELIMSSVKAQFGHCEVASGLI 1140
QY 120 Q-----Q-----Q----- 122
Db 1141 QLMKVSSIGKHGIIPSVHNILPSEHNRNENIRLPFAEEKQIDRSATVSFGITGKTIV 1200
QY 123 -----SAK----- 125
Db 1201 VITERVSQLNVDNIQNCYLLPVSAKTKDGLKKACLSLIE MIDNSCESLYDITTLQOKQT 1260
QY 126 N-----G-----E-NT--AN-----G-----E----- 134
Db 1261 NPKWRTAVVSGSHADVVLKQFLTSEHNTSLTNMHI STSHSIGCTFFHNIPEFDHY 1320

QY 135 -----E-----NG-----A-----H-T-----IA-N----- 143
Db 1321 SMFCHELRPEPHSNTNSIYHLLAVVYALRVILKHLKLTNSFAVGGSLSLVLAAVDAAPS 1380
QY 144 -----N--HT--D--MM-----E-----V 151
Db 1381 HYLNDLLHAFANDDVRMMKRIARDVTISIE NVKLLNAGEPIITTAQAVEATIDQKVKNV 1440
QY 152 -----D-----G-D-----V--E-- 156
Db 1441 RLPEPTLILSPSRAVEFASQLETIQDYKLGKFTQGGQVDPAGIFGTPIKILDILPEYP 1500
QY 157 -----IP--P--N-----KA-----VVL-----R- 166
Db 1501 FNRKSPWLPIPDSPVSNKGEKPLIPKSYEFLLKSKQWQHVNHVDSKIVLPGATSIRL 1560
QY 167 -----G----- 167
Db 1561 VHQLANGKPTVELSNIDFLNKITPSEAPSVVKIEODGLEKLVFGETDAISFKLTQLQNFN 1620
QY 168 -----H-----E-----S-----EV-F----- 173
Db 1621 PIPNEKLNABVHHTNIYERFANSHLTYNREFQMVDSLKYTMKGGEVRFVSMKDLIDIL 1680
QY 174 -----I--C----- 175
Db 1681 GTLQALVGCYFPENTNDNSPFVFTIDQLSILNGDISQKQLHAVLYKVDSSGNPINGDATV 1740
QY 176 -----A----- 176
Db 1741 YDALGNIILHISNVTFKRLNGQASPLSTKTVDSKTIKKVENEODKRASKMHLWVFE 1800
QY 177 -----W-----N-P-----V-----S-DILASG----- 187
Db 1801 NFGWTDIDNTTGFDFDLGLTSIOAVKLRNAIKSNYPNASSTCVFDPYPSIDLL-SGYLSTLN 1859
QY 188 -----S-G-----D-S-----T-----A-R-----I-----W----- 196
Db 1860 DPQVTSIGEDDIQKDLTEDHKPFLAENPIGVMAACRLPGGVSSPSSELMLLKGKN 1919
QY 197 -----N-L-----S----- 199
Db 1920 ASSRIPATRPVTRNTLISGSKYGNPVEGGNFITQDVTPDFPFKISKSEALIDPQORL 1979
QY 200 -----ENS--TS-----GS----- 206
Db 1980 LLECVOECLENSQVETSNVGVFVGLMEKEYQDMMESSILAMLGSMMAAIAGRVNYIFG 2039
QY 207 -----T----- 208
Db 2040 CYGPSVTIDTACSSSLVALEMAINALLDNRC SKVIVAGVNLILNEKGQGLRTNGKLSQH 2099
QY 209 -----L-----VL-----R----- 212
Db 2100 GMSLSPDSRASGYGRSGCVLMLELAKPNFHMSTIQSVNVNHHGGRSVSLTAPNVAHK 2159
QY 213 -----H----- 213
Db 2160 MLTTSVINGQSPSLAIDYWEAHGTGTPGLDPIEBFTLSSILQNIIGSVKASLHGGEASAG 2219
QY 214 -C-----IR-----E-----G-G-- 219
Db 2220 TCGLLKLFLMLTYQVPTLIHFHVLNKOINAGSIRLPIIGEDSELVSAGISGFSVSGTNA 2279
QY 220 -----Q-----DVP----- 223
Db 2280 AAIAFNDNNKLEPIPIHKYYILPISAKNQISLDNLEKQILSVIPLTDVPICNIASALAN 2339
QY 224 -----SN-----KD----- 227
Db 2340 NRSHEFTIRNALIVSNSGIVNSKMEKPHRAKDRYHVKLCDLSDASLLQYDVINETYT 2399

QY 228 V-----TSL-----D-----W--N-S-E-- 236
Db 2400 VASLKNQSFAMKFAIKFSLTSLSEVIEIVASDGBELLAVLANGSLKWFNFKTWIELP 2459
QY 237 -GTL--A-----TG-----S 244
Db 2460 IGSLLTEFADHLNLTSSIKSYQTPSHNLDSPMELMKLMKIYITGYDWDATVYS 2519
QY 245 -----Y-----D-----G-----F-- 248
Db 2520 PVEQFIALPNYQFNKQTLWFERLEIVDHYLIGTIDESEDTLILKNQISELHPQFPFKG 2579
QY 249 -----A--R-----I-----W--T----- 253
Db 2580 KPLDVCTMSEIAIEALKIRNEIPFSIQNLKTELITLTKPAWLETNVTNRNEDDEGFNVSA 2639
QY 254 -----K----- 254
Db 2640 IDQRLFSNLASSVEIQIEVPAVEQIPDKVVYLKECPNAVIRRHNNMVYDSRAEQSP 2699
QY 255 -----DG-----N-- 257
Db 2700 FRTANIVLNEIIGFAPTPSDMFIEILGVLPVSHVMVQVDDGALWQPMISQDKRVLNITY 2759
QY 258 -----LA----- 259
Db 2760 VLKDAKGLPIPTIRMHKSTLLSSQASIVAAKTLQMAVRHKVCLAVGDVIESGLDIDES 2819
QY 260 --ST--LG----- 263
Db 2820 QLSGTSELGIDSLATVOLLNRLNKQYFPEIELTTSDFDNPSIIDLSIMIBQLLNKEGI 2879
QY 264 -----Q-----H-- 265
Db 2880 TEPSEPNTPKTSLRGRKLSIPAVRAQVLAQIEFVENYNSKQEVQAEAPSSSECSNHLE 2939
QY 266 -----KG-----P 268
Db 2940 ESDATVDRTEIRRKVSLAVFDLATETLSAEDLQSKGFTLGMDSLSIVDFVRLNDKYFP 2999
QY 269 -----IF-----A----- 271
Db 3000 DDEITASDIFDYPTVDELSDHIVRKSSVPPAASEIMKETMNGISTSVDAEHTKLENLS 3059
QY 272 -----LK--W--N--K--K--CNFI--LSA-----G----- 285
Db 3060 QSPMLLENQNSINPTLKMINSQTIKLVKPSDGNFLFELNANGQGEKEIQKHTGPNNII 3119
QY 286 VD-K-----T-T-----I----- 291
Db 3120 IDLKGFEHGSTETLYMSLLNLVKSISKLEIQCFGVSGFGLGNSISRAFMKTVAAEKNP 3179
QY 292 -I--W-----D----- 294
Db 3180 LISFAWQNVQVSVFVDSPIGTGNWLTGGLSGIGLEIGKFIANGNAENVILISRRQPT 3239
QY 295 --A-----HTG----- 298
Db 3240 AKALRDILSTELTHIGLARTIVILKIKNISAKLIVQSFKLSFSTKYKVISDLFSKKVTF 3299
QY 299 -----E----- 299
Db 3300 FYNLQSCFCKHFWLFEWLEVYLKQWTVHTIAADINDKEKILRELKLNVTGIIH 3359
QY 300 -A-----K--Q-----Q-P-P-----FH----- 307
Db 3360 SAGVLKSKIERQKESFNQVFTPKANGFHVLEIEKHFNYKIENFIMMSFTAACNEG 3419
QY 308 -----S----- 308
Db 3420 QLNYSVNAYLEYQVORRRRQSGCAIQWGNWIDTGMATDENVRKFLANLGLFOHKNKD 3479
QY 309 -----A-----PAL-----DVDW-----Q-----S 318

Db 3480 ALKYLRACILTKPELIMVANDMNVLKNRDLPKDLINTGILPFDFTGKNESEFPPLS 3539
QY 319 N----- 319
Db 3540 NGDFEKVSMNFSVEDEEBEVLLEIKVKSIIILCMSPTKLNNKNIMDMGLDSKLIIVEFLNF 3599
QY 320 -N-TF-----A-----S--C----- 325
Db 3600 INSTFKISUNLSDAVNHTPLEKLAHIPEQMTIVDHPVNSVKSEIFKSTDCPIFGINI 3659
QY 326 -----ST-----D----- 328
Db 3660 PFDKNDFDKAKSTAVKLLNGEQLPTAGKYAVSVVGKSIIRDVWSKIEAAPQIKLCOE 3719
QY 329 -----M-----C-----I----- 331
Db 3720 SSKCVMLTGOQSYPMGRQLVENEYIFRTTLOSLCKKCDYLOQDVLSLWELFNTHY 3779
QY 332 -----H--V--C----- 334
Db 3780 KLLQUTKMQPIMFCGYATAQLMISLGIYPDYILGSHVSELVAGVLAGIMSIEDGLRLI 3839
QY 335 -----K-----LG-----Q-----D-----R-----P--I--KT-----F- 345
Db 3840 VERGKAMENIAGLGALLAVQREIADVLKRPKVSIVATINSKQVVFAGTKSVLDAALAEV 3899
QY 346 --QG-----HTN--E-----VN--A 354
Db 3900 KQGGKQATVYNOQYPPHNSLIQETHLVSLRQCLADIKFSAGRTPLVSNVTGOIINTFSEA 3959
QY 355 -I-----K-----W-D-----PT----- 360
Db 3960 YIVKHTVSAKVDCVETLOAKGVTVWIDAGSAVLATFVKRIIQPTLSKRIIVQTCKE 4019
QY 361 -----GNLL-A-----S-----C-----S----- 368
Db 4020 KESDVLNVQACLELEQSGLPISWTLYGCGRNADERLVEFPMTHTNDIINKDEFELLEGH 4079
QY 369 -----D-----D-----DMT-- 372
Db 4080 QLNKIVVAGAYQLFKIDQLVKLKAAGMELMLKNVKFLKPWYIEDNREYQIOWNSDMTIE 4139
QY 373 -----LK----- 374
Db 4140 LIVNSVIVCSLEVEPQNSVLKLETISENEKEPVEHDFYETLFRNGIQLYDSGFRRIESARR 4199
QY 375 -----I-----W-----S--MK----- 379
Db 4200 SDRKCFSQIKSSPFAWPLIDSAMHSITASVVRPRDCYFLPVAMGSVMTKDTNSFTLPNL 4259
QY 380 --Q-----D--N-----C-----VH----- 385
Db 4260 HAQTVITSETDKFIQNVNALLAGDTPICEVRNMTIVLTKTEPVTHTIPNSIETVETPPKS 4319
QY 386 -----D--L--O-----OH-----N--KE--I----- 394
Db 4320 EIEIVGFDISLPYQISENSENWQHLKNTYVQKLNHNSLKQDHARVALLDSADRYWDPPE 4379
QY 395 Y-----T-----I-----K 398
Db 4380 YFGIRPSEAKFIDPOORLLILCSVAKLLDSLLITSITSNTGVFIGCSANBFHIVVAYGK 4439
QY 399 -----WS-----P-----TG----- 403
Db 4440 DPRAEMSGTNSALAGRIAHWLKLGPPVVTLDTACSSSFYALSAACDALRTGQCEYAI 4499
QY 404 ----- 404
Db 4500 GTVNLVMEHTTDLQNAKQVTDDFCFADFVANGYKSEAVCSMLLTKSPNIDSVATIT 4559
QY 405 -----GT-----N-----NP-----NA----- 411

Db 4560 NYATGHNGTSSLSFTPNGLSOLVEMORATNPLEKILEIOTHCTGTGLGDPFIEINAIKSLV 4619
QY 412 -----N-----LM-----K 414
Db 4620 SSACKIGSVKSNIGHTEGSGLVSLCSSLSFRSKYRVAQLHLKCPNTSINKTWMICRFI 4679
QY 415 -----L-----A-----SA-S----- 419
Db 4680 GEDADENNSILINNFPGTSGNCVWLKPKNAISEHFVSSEVFPYILLSSHSAKSLQKYVQ 4739
QY 420 -----D-----S-----T-----VR----- 425
Db 4740 VLCEFISNSAKSLHDMMSLFQKKIHVHRQFIIPNFKRIAVTSLDGFVEVRDERLEKX 4799
QY 426 -----L-W-----D-VDR----- 431
Db 4800 HPCSFLKEGVVHFDKDKFORVDLPSIVNNTLHWALDSYRDEIDRHESQMSFKNIFY 4859
QY 432 -----GI----- 433
Db 4860 EKVLETPMPQONPIKVVVCIGRLDLIPKIEIDVSSFANGIIIVHPISNISFEYLKLA 4919
QY 434 -----CI-----HT-----L-----T-----K-----H- 441
Db 4920 VMSLISRNQNVFIIICFENGTSHTMTGTLRSLASEKMIPIYKFVSDIKVDALKLJEFNHE 4979
QY 442 -----Q----- 445
Db 4980 YMFELFYKRGYVERLKVNTFPKAPQYKELISGGTGGIGSAIINELKPKSSVIIT 5039
QY 446 -----Y----- 446
Db 5040 RKNIASEDGKTPLSSDITRLDISHKFNYPVFLAGIVNNSLHENVKRDSDLDMVSIKLOGA 5099
QY 447 -----S-V-----AFS-----P-D- 453
Db 5100 KNLKCCDETHFVFSSTIANVLGSGYSQSAFNSGLVTSFLETSKSTIIHWGPWKDV 5159
QY 454 G-----R-----Y-L-----A-S-----G----- 460
Db 5160 GMLAQPERREIVKQIESNGKWLPNQDALSIVFTQFMETHQIIVFDGDFDTIVARQPHL 5219
QY 461 -----S-----F-----D-----K-----C----- 465
Db 5220 OKLLSEVVEKTKYKEIKKSLNFEEIPEIVGITDISKLNIPFMDLGLDSLCMENLRY 5279
QY 466 -----V-----H----- 467
Db 5280 SLNKNFDLELTVSEMFENATYQKLOTYVETLRAKHSLSLSDRVSSQVSNKEDDTRVAV 5339
QY 468 I-W-----NT-----Q-----VC-----L----- 475
Db 5340 IGWSAEFGSSNIHEYENLMDGICSTGNKYLKKNPFGDNKFNLTDEARVLPQVR 5399
QY 476 -----H-YL-----N-G-----Q----- 481
Db 5400 KFIQHAYLALENSYYKQHELRCGVFAGAPSDYGRADDDHDMKFLFMNNSYLASY 5459
QY 482 -----VL-----L-----N-----L----- 486
Db 5460 ASYCLDLKGEAVSVGSACSTALVAVANAVKSIQSGSMYALVGAASIAEVSGALSGFDDQ 5519
QY 487 -----G-R-----S-I-----C-L-----YT-----L----- 495
Db 5520 KKTMFSGVCRPFDKDSGIVRGSGVGVFLKRYSQLLDNDNVHFVKDFAINDGHS 5579
QY 496 -----P-----H----- 497
Db 5580 RASFMAPAGQLKCMVDVLARFTNKEKERSIFVECHATGTTLGTIEMNSLRTAYSFKN 5639
QY 498 -----H-----LV-----VIP----- 503
Db 5640 KLAIGCKANIGHAYAAAGLAALVKCAKMLQGTIIIPQVNFSEFRDGMGQFFTVNGKST 5699

QY 504 -----LVAL-----IEL-----LVL-----K 514
Db 5700 ISQNSLISIDSPGIGGTNNVHMVIEFPARQSEVVVKISSENILIDYDMIPISAK 5750

RESULT 8

T13931

projectin - fruit fly (Drosophila melanogaster) (fragment)

C:Species: Drosophila melanogaster

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C:Accession: T13931

R:Daley, J.; Southgate, R.; Ayme-Southgate, A.

J. Mol. Biol. 279, 201-210, 1998

A:Title: Structure of the Drosophila projectin protein: isoforms and implication for pro-

A:Reference number: Z17815; MUID:98300339; PMID:9636710

A:Accession: T13931

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-6658 <DAL>

A:Cross-references: UNIPROT:O76281; EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AAC3

C:Genetics:

A:Gene: projectin

A:Cross-references: FlyBase:FBgn0005666

A:Map position: 4

A>Note: intron positions not resolved (incomplete sequence)

C:Keywords: muscle

Query Match 70.5%; Score 2572.1; DB 2; Length 6658;
Best Local Similarity 9.1%; Pred. No. 2.4e-52;
Matches 438; Conservative 66; Mismatches 5; Indels 4309; Gaps 375;

QY 1 MSI-----S-----S-----D-----E-----V-----N-----F----- 10

Db 20 KSLIAKRFVKPFIVGEGLKNTVVKKGITRFDIKYDGEPEPAATVWKTGDNLFKFNQRI 79

QY 11 ----L-----V-----Y-----L-----Q-----ESG-F----- 21

Db 80 CLDQLERNSSITIKSVKDKTKYKLVLSNSSGTIESEAQVVVLDRLPLPGGPFEEIR 139

QY 22 -SH-----S-----A-----F-----T-----F 28

Db 140 ASHIKMKWKRPDDGGCEISGVALERMDEBTGRTWIPAGEVGNPNETSFDFKGLTPNKKYKF 199

QY 29 -----Q-----I-----KS----- 32

Db 200 RVKAINKEGESEPLETFDASCKNPYDPPSPQPVIDDYDNKSVLLKWKRPSPDCGRPT 259

QY 33 H-----I-----SQ----- 36

Db 260 HYIVELKDKFAPSWSEVAKTDDPNPECNVEGLKEMVYQFRVAVNKNAGPSEPSQPTDNH 319

QY 37 -----S----- 37

Db 320 LCKHKNLKPDIDRSTFKRVTIKSGRTHKWSVDLGEPIPELHWSWRDDIPLTNGDRIKIE 379

QY 38 N-----I-----NG-----A-----A-- 42

Db 380 NVDYHTDFSITNVLKROSGFYTLKAENRNGIDRETVELVLGKPKSPKGLAVSDVTASG 439

QY 43 --L-----V-----PP----- 46

Db 440 SKLQWKKPDDVGVPIKEVVEKMDTATGKVRVGRSPGCEKPEPFDVTGLSLGSEYMPR 499

QY 47 --A-----A-----LI-----SI-----IQ 54

Db 500 VSAVNEGEAEPLTTLVGVAKDPPEPNKPGTPEVTDYDNQISLSLKAAPANNDDGAPIQ 559

QY 55 K-----GLQ----- 58

Db 560 KYIIEKKNKNTWEKALEIPGDQLEATVAGLQYGEYQFRVIANKAGLSPPSDASVPQ 619

QY 59 -----Y-----V-----E-A-----E 63

Db	620	IVKYKLLKPRIDRSLNKLPLLRAGKPIRYDVNVRGEPAPVITWYQNDKELKPEELPSSSE	679	Db	1700	LVTDIPTKAKNPPDAADTPGKPIQIVDSGNHCDLKWRAPEDDGGASITGYIVERKDPNTG	1759
Qy	64	-----VSI-----	71	Qy	190	-----D-----S-----TA-----	193
Db	680	IKNPYNTKISIIETVRKHTGIYKIIAVNEHGQDEATVEVNILAPPSKRGPLDVKDVTK	739	Db	1760	KWQKALETSTDPCKARVNDLIAGNKYQFRIMAVNKAQKSPSESDQMTAKDRFAPPKID	1819
Qy	72	-----L-----FD-----GR-----P-----	77	Qy	194	-----R-----I-----W-----N-----LSEN-----S-----T-----	203
Db	740	DSCKFKWKPPEDDGGKPIGSAIVKEKFKQGRWVPLGRISANDTEFDVKGLOEGHEYQFR	799	Db	1820	RTNIDITSKAQHIFRFDIKVSGEPATKWLHNKARL-ENDDSNYNIDMESYRTKLTVP	1878
Qy	78	-----I-----ES-----L-----SLI-----DA-----V-----M-----	88	Qy	204	-----SGS-----T-----Q-----LV-----LR-----H-----C-----	214
Db	800	VKAINEGESDPLDSDSIIAKNPYDAASKPGTFNIVDNEHMKLWEAPRSDGGAPIS	859	Db	1879	ISKRFHSGKYTLKAENESGRDEASFEVILDKPGPEGLRVTDVHKGCKLKNAPLDD	1938
Qy	89	-----P-----D-----V-----V-----Q-----T-----	94	Qy	215	-----I-----R-----R-----R-----R-----R-----EG-----G-----	219
Db	860	GYIIKKDKFSPIDWIBILSTNTSVPEATVEGLVEGNIYQFRVAVNKAQSGFSDPDATEPH	919	Db	1939	GGLPIDHYIIKMDVSGRWLPSGRFSPFAELNNLEPSHEYKFRVLAVENTGESEPLTG	1998
Qy	95	-----R-----Q-----Q-----A-----Y-----R-----	100	Qy	220	-----Q-----	221
Db	920	LAKPRNLKPYINRDKMKPIKVRAGQPVKFDVKGEPAPSLTWFLKETELTSTGQVRLEN	979	Db	1999	EQSVIAKNPFDGPKGTPEAVDWDKHDVLDVWRPPIINDGGSPITGYVVEKREKGTDMI	2058
Qy	101	-----D-----KLA-----Q-----QA-----A-----	109	Qy	222	-----VPS-----N-----K-----D-----VT-----	229
Db	980	IDYNTKLTLLDTRKQSGQYKLAENINGVDEAVVEVILDKPSKPEGLEVSIDIHKEGC	1039	Db	2059	KGTEITIPCLGEBEKATVPTLNENCEYEFVRVKAINAAGPEPSDASKPIITKPKLAPTI	2118
Qy	110	-----A-----A-----A-----A-----A-----	112	Qy	230	-----S-----LD-----WN-----S-----	235
Db	1040	KLWRKPKDDGGIPITGYVIEKMDTATGKWVPAGSDPEKYDIEIKGLDPNHRHYQFRVKA	1099	Db	2119	LDPKNIYNTYKSGEPFLDINISGEPADPTVWQNNKSVQTTFSHIENLNYNTKYIN	2178
Qy	113	-----A-----A-----AA-----	116	Qy	236	-----E-----G-----T-----	238
Db	1100	VNEGESEPLETESATAKNPPDVSAAPGLPELEDWDEHHVKLWBPPIRDGSGSPITNYI	1159	Db	2179	NNPERKDTGLYKISAHNFYQDQVBFQINIITKPKPGGLEVEVHKGCKLKWKKPKD	2238
Qy	117	-----AA-----S-----QO-----G-----S-----AK	125	Qy	239	-----LLA-----TC-----S-----Y-----DG-----FAR-----	250
Db	1160	IEVMDKDPGEFVRAVETDSPVCKGVVKLEGGQYKFRVAVNKAQSGPSEQTNNHVAK	1219	Db	2239	DGGEVPSYLVKFPDPTGIMLPVGRSDGPEYNDGLVPGHDYKF-RVAVNKGESSEPL	2297
Qy	126	-----N-----N-----GE-----	128	Qy	251	-----I-----WT-----KD-----	255
Db	1220	PRELKHIDRWNLKPVIVTGLSISLDINIRGEPAPKVEWFFNNSSVTSDEHSVKIDNVD	1279	Db	2298	ETLSIIAKDPFSVTKGCVPEPTDWTANKVELAWPEPASDGGSPIQGYIVEKDKYSPL	2357
Qy	129	-----NT-----A-----N-----GE-----E-----N-----	136	Qy	256	-----G-----G-----G-----G-----G-----A-----	259
Db	1280	YNTKFFVMAQRSQSGKYIIKATNEVEGEDEAELEVTILGPKPKGFLQVNDITKHSCKL	1339	Db	2358	WEKALETNSPTTATVQGLIEGNEYQFRVAVNKNLSEPSDPSKIIFTAKKRYIAPKIDR	2417
Qy	137	-----CA-----HT-----I-----A-----N-----	143	Qy	260	-----S-----T-----L-----GQ-----	264
Db	1340	KWEKPDGSGSPIDYVEIEKLDPHTGQWLPCKGKSTPEAKVIGLHEGKAYKFRVAVNKE	1399	Db	2418	RNLNITLSSGTALKLDANITGQAPKVEWKLSNYHLSQSKNVTIETPDYTKLVIRPQ	2477
Qy	144	-----N-----N-----HT-----DM-----M-----	149	Qy	265	-----HK-----G-----	267
Db	1400	GESEDLTEKPIIAKNPYDEPRGKPEPTNWMDKFDVLDLAWDPKNDGGAPIQYIOMR	1459	Db	2478	RTDSGEYLVTAINTSKDSVLNVNVTIDKPSPPNGPLQISDVHKEGCHLKWKRPSHDGT	2537
Qy	150	-----E-----VD-----GD-----VE-----I-----P-----P-----NKAVV-----	164	Qy	268	PI-----FA-----LK-----W-----NK-----K-----G-----	278
Db	1460	DKSGRAWDSATVPGDKCNGTGTVEEGHEYEFRIVAVNKAQSGSDSDYSKSVIAPRFL	1519	Db	2538	PIEYFOIDKLEPBTGOWIPSCRSTAPQVDVTGLSPGNEYKFRVSAVNARGESQPLVDES	2597
Qy	165	-----L-----R-----G-----H-----E-----SEV-----F	173	Qy	279	-----N-----F-----I-----LS-----A-----G-----	285
Db	1520	KPHIDRNLOKIMRSQMLHIDALIKAEPPAKVTWYNTKTEIKTSDHIKIENEDYKTF	1579	Db	2598	IVARNPDEFGKPNLKAIDWDKHDVLDLAWTPPIDGSGPISCYIEKQDKYKWERALD	2657
Qy	174	-----I-----C-----A-----	176	Qy	286	V-----D-----K-----T-----T-----T-----II-----	292
Db	1580	IMPKVKRADRGIVYIVTAKNDGSGTVEVEVLCKPSKPGPLAVSNVTAETHLKWKEP	1639	Db	2658	VPADQCKATIPDLVEGQTYKFRVSAVNAAGTGEPSDSTPIIAKARNKPPIIDRSSLVEV	2717
Qy	177	-----W-----N-----	179	Qy	293	-----W-----	293
Db	1640	EDDGGDPIEQLYVERMDTETGRWVPVLTITKTPPADVTGLTEGKEYLFRVAVNSEGESEP	1699	Db	2718	RIKAGOSTFTDCKVSGEPAPQTKMLKKKEVYSKONVKVTNVDYNTKLKVNSTRSDSGI	2777
Qy	180	VSDL-----L-----AS-----G-----SG	189	Qy	294	-----D-----A-----H-----	296
		-----L-----		Db	2778	YTVFAENANGEDSADVKKVTVIDKPPAPNGPLKVDEINSESCILHWNPPDDGGQPIDNYV	2837

QY 297 -----T-GE-----AK-Q-----Q----- 303
Db 2838 VGKLDDETTGRWMTAGETDGPVTKLVGGLTPGHKYKFRVRAKNRQGTSEPLTTAQAIK 2897
QY 304 -----F-----P-----F----- 306
Db 2898 NPPDVPTKPGTPTIKDFKDFVLEWTRPEADGGSPITGVVVEKDKFSPDWEKCAEISD 2957
QY 307 -----H-----S-----A-----P----- 310
Db 2958 DITIAHVPLDIEGLKYEFVRVAVNAKAGPSPDATETHVAPKNTPPKIDRNFMSEDIK 3017
QY 311 A-----LDV-----DW-QSN-NT-----F----- 322
Db 3018 AGNVFEDVPVTGEPLPSKDWTHEGNMIINTDRVKISNFDRTKIRILSATSDTGVYTLT 3077
QY 323 -----A-S-C-----S-T----- 327
Db 3078 ARNINGTDRHNKVTILDAPSVPEPALRNGDVSKNSIVLRWRPPKDDGGSEITHYVVERM 3137
QY 328 D--M-----C-----I-H-----VCK-----L----- 336
Db 3138 DNEAMRWVPVGDCTDEIRADNLNENHDSFRVAVNAKQSQPLTTSQPIAKOPYSH 3197
QY 337 --GQ-----D-----R-----PI-----KTF-Q-----G-----H 348
Db 3198 DKPGQOATDWGKHGFVLEWSTPKRDGGAPISYIIEKRPK-FGQWERAADVLDGNCKAH 3256
QY 349 --TN-----E-----VN-----AI-K----- 356
Db 3257 VPELTNGGEYEFVRVAVNAKGGSPSPSSITICKPRFLAPFDKSLNDITVHAGNLG 3316
QY 357 W--D-----P-T-----G-N-----L-L----- 364
Db 3317 WTLPIBASPRPLITWLYNGKEIGSNRSGSLFQNELTPEIVSSLSASRGYTLILKNEH 3376
QY 365 --AS-----S-----C-----S-----D----- 369
Db 3377 GSFDAHAHATVDRPSPKPLDITKTRDGGHLTWNVPDDGGSPILHYIIEKMDLSRS 3436
QY 370 --D-M-----T-----L-----L-----K 374
Db 3437 TWSADAGMSTHIVHDVTRLVHRKEYLFRVAVNAIGESDPLEAVNTIIAKNEFEDPDAPGK 3496
QY 375 -----I--W-----S-----MK----- 379
Db 3497 LIITDWRDHDILQWAVPKSDGAPISEYIIQKKEGSPYTNVVRHVPNSKNNTTIPILT 3556
QY 380 -----Q-----D-----N-----C-V-H-D----- 386
Db 3557 EGQVEYEFVRVAVNAKQSPSPSPDMIMKPRYLPPKIITPLNEVRKICGLIFHTDIHI 3616
QY 387 -----L-----L-----QO-----H----- 390
Db 3617 GEPAPATWTLNSPLLSNDRSTITSIGHHSVVHTVNCQSDSGIYHLLLRNSSGIDEGS 3676
QY 391 -----N-----N-----K-----EI-YTT-K-----W----- 399
Db 3677 FELVLDKRPPEGPMMEYEITANSVTISWKPCKONGSGSEISSYVIEKRDLTHTGGGWPA 3736
QY 400 -----S-----P-----S-----T-G-PG----- 405
Db 3737 VNVSAKNHNAVPRLLGTYELRVMAENLQGRSDPLTSDQPVVAKSQYTVPGAGKPE 3796
QY 406 -----T-----N-----N-----P----- 409
Db 3797 LTDSQKNHITIKWQPIISNGGSPGIIIGYDIERRDVTNGRWIKNGQVPVTAEOYDDRVTSN 3856
QY 410 -----NA-----N-----L-M-----L-----A----- 416
Db 3857 HOYQVIRISAVNAAGNKTSEPSAIFNARFLREKPRFYFDGLIGKRIKRVAGFVNINPI 3916

QY 417 S-A-----S-----F-----DS----- 422
Db 3917 SGAPTPIEWKRGDLKEGKRISYETNSERTLFRIDDSNRDRSGKYTVTAANFCKOTA 3976
QY 423 -----TV-----R----- 422
Db 3977 DIEVIVDKPSPPEGLSYTETAPDHISLHWYSPKDDGSDITGYIIEFTFGVDDMKPV 4036
QY 423 -----L-W-----DVD-----RG-----I-C----- 425
Db 4037 PGTCPNTNFTVKNLVEGKKYVFRIRAEINYGASEALEGKPVLVKSPDPGAPSQPTISA 4096
QY 426 -----L-W-----DVD-----RG-----I-C----- 434
Db 4097 YTPNSANLEWHPPD-DCGKPKITGYIVERRERGGEWIKNNYPTPNTSVTVSNLRDARY 4155
QY 435 -----I-----H-----T--LT-KHO-----EP-----VYS-----VAFS-----P-DG-454
Db 4156 EFRVLAVNEAGPGHPSKPSDPMTHAQRYRPPPEPKPDRI-TRNGVTLSMRPPRTDGK 4214
QY 455 --R--YLA-----S-G--SF-----D----- 463
Db 4215 SRIKGYVEMRPNGKDWKTVDIPINSTVTVVPSLKEGEYSFRVVAENEVGRSDPSKP 4274
QY 464 -----K-----C-----V----- 466
Db 4275 SQPITIEQPNKPCMELGKVRDIVCRAGDDFSIHVYLPAPKPNAPFWSNDNMDNNRV 4334
QY 467 H-----I----- 468
Db 4335 HKHLTDAASVVVVKNSKRDSQYRLQFKDTSQFDTATINVRVLDPRSPPTRLRADVFSG 4394
QY 469 -----WN-----T-Q-----V--C----- 474
Db 4395 DSLTYWNPDDGGSAIQNYIIEKKEARSSTWSKSSECTVLVFRIRNLVNLKEYDFRV 4454
QY 475 -----L-----H----- 476
Db 4455 IAENKYGQSDPANTSEPIILARHPDIPNTPGPHGIDSTEDSTIIATWKPKHGGSPITG 4514
QY 477 Y-----L-----N-GO----- 481
Db 4515 YIIEKRLSDDKWTKAVHALCPDLCKIPNLNENAEYEFRAVNAAGOSAYSGSSDLIF 4574
QY 482 -----V-L-----LN-L-----G----- 487
Db 4575 CRPPHAPKITSDDLIPDMTVIAGDEFRITVPIHASPRTASWSLNGLEVIPEERIKFDS 4634
QY 488 -----R--S-I-----C-----L--Y----- 493
Db 4635 NDYASMYNKSARKDETSYITLTNNKGSDDTASCHVTVDRLPPQGPLNAYDITPDTIC 4694
QY 494 TL-----P-----H-----HL-----V----- 500
Db 4695 TLAWKTPLDGGSPITNYVVEKLDNSGWSWKISSFVRNTHYDVMGLEPHYKYNFRVAEN 4754
QY 501 -----VI-PLVA-----L----- 509
Db 4755 QYGLSDPLDIEFPMVAKHQFTVPDEPGQPKVIDWDSGNVTLIWTRPLSGSGSRIQGYIE 4814
QY 510 -----LVLK 514
Db 4815 YRDILNDSSWNAVDYIIK 4832

RESULT 9

A45086

HC-toxin synthetase - fungus (Cochliobolus carbonum)

C:Species: Cochliobolus carbonum

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Nov-2000

C:Accession: A45086

R:Scott-Craig, J.S.; Panaccione, D.G.; Pocard, J.A.; Walton, J.D.

J. Biol. Chem. 267, 26044-26049, 1992

A;Title: The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous
 A;Reference number: A45086; MUID:93100328; PMID:1281482

A;Accession: A45086

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: nucleic acid

A;Residues: 1-5232 <SCOs>

A;Experimental source: SB111

A;Note: sequence extracted from NCBI backbone (NCBIP:120884)

C;Superfamily: Cochliobolus carbonum HC-toxin synthetase; acetate-CoA ligase homology; a

F;277-743/Domain: carrier protein; phosphopantetheine; phosphoprotein

F;784-854/Domain: acetate-CoA ligase homology <ACLI>

F;1882-2366/Domain: acyl carrier protein homology <ACP1>

F;2397-2464/Domain: acetate-CoA ligase homology <ACL2>

F;3036-3526/Domain: acyl carrier protein homology <ACP2>

F;3349-3618/Domain: acetate-CoA ligase homology <ACL3>

F;4186-4655/Domain: acyl carrier protein homology <ACP3>

F;4680-4750/Domain: acetate-CoA ligase homology <ACL4>

F;818,3583,4715/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 70.48; Score 2568.8; DB 2; Length 5232;

Best Local Similarity 8.9%; Pred. No. 1.8e-52;

Matches 447; Conservative 54; Mismatches 6; Indels 4542; Gaps 379;

1 MS-----I-SS--DE--V-NF-----LVY-----13

179 MSAMEVKNVSTLTLYIMSSDPDKTAINRLSIPRDLAQIMRWNRDLKKSERTNLVYDLFS 238

14 R-Y-----R-Y-----LQ-----E-----S-19

239 ARAHEQDANMAIDAWDGRMSYTELERVSTWARQLQKQIGISQGSWVLFCEKSRSLAVSM 298

20 -----G-----F-S-----22

299 IAILKAGVCVPIPRVPERIRDIIRTNATIALVAGKTAALFKSADTAVOTIDITKD 358

23 --H-----S-----AF-TF-----GI-K--SH-----33

359 IPHGLSTVVQSNKTDPAFGLFTSGTCVPRKIVVTHSQICTAVQAYKDRGVTSER 418

34 -----IS-----Q-----36

419 VLOFSSYTFDISADTFALFYGGTLCIPSEEDRMNLQDMVSVRPNMAVLPTVSRLF 478

37 -----SNI 39

479 DPGVVDFTSLTFTGEASREADVPIEAGVNLVNYGPAENTLITTRIRKGSNI 538

40 --N-----GA-LVP-----P--AA--L--I--50

539 GYGVNTRTWTVDSGACLVFVGSIGELLIESGHLADKYLNRDPTTEAFLSDLPWIPNVE 598

51 --SI-----I-----Q-K--GLQ--Y--VEA-----E 63

599 GDSVRRGRFRYRGDLVRYCDDGLICVGRSDTQIKLAG-QRVELGDEVAHLQSDPTTSQ 657

64 --V--S-----I-----NEDGT-----L--F-----73

658 AAVFPRSGPLEARLIALLVGKNKGTPHNNQSLPKPAFAQCPPDLVKVATSLQORLP 717

74 -----D--G--R-----P--I--E-----79

718 YMVPSVWLIGDPLPMSVSGKLDRAVLQDQLESLSPSDYAEILGTTGLEVDPGCAASSVAS 777

80 -----SL-----SLI--D--A--V--M-----88

778 DSDLRMDNDDSLLTACSRVNLNPAKISYSQSFIHAGGDSITAMQVSSNMKFTGKRIG 837

89 -----P-----D--V--V--QTR-----Q--A--Y--99

838 VKDLLVSPSISTAACTKASQDSRNFVAPQO-RIPVSPQKLFQTAESKSNHYH 896

100 -----R-D--K-----L-AQ-----Q-----AAA-----110

897 QSFLRIDQPIKQPIQTIEDAISLVQRHPLQARFERTEGDWYQYIPIDVERASVEVIG 956

111 -----AA--A-----A-----A-----115

957 SLSTDREAAMLRAROSIDLTEGLPLRCQLFNNNVDEASRLFPFWIHHAVIDLVSWRIIM 1016

116 --A--A--A--S-----Q-CGSA-KN-----G 127

1017 ELELAHLADTSTPDGRGEAVOESVFPLAWCQVQAEAKVIPDRTVPLIPKIPTADFGYWG 1076

128 -----ENT-AN-----G--E-----E-----135

1077 LKHDENVYNTVERKIPGLGHSITEDLLYKCHDSLHTKTIDVLLAAVLVSPRRSFLDRPVP 1136

136 --N-GAH-----TI-----A--NN 144

1137 AVFNEG-HGREPGGEDAVDLSTRTVGWFTTISPVYVEVSPGDILDVVRVKDYRWATPNN 1195

145 -----H-----T-----D-147

1196 GPDYFTKYLTSQGIKLFEDHLPAPILFNVEGRYQAMESEQTVLKPSWHAGEASKDDDP 1255

148 -M--M-E--V--DG-----D-----V-EI-----P 158

1256 GLRRFCLFEISTAVLPDGLHLTCMNKMRHQGRIRLMDLTLLPAAIGEVSSLASASP 1315

159 -----PNKAV-----VLRG-H-----ES-170

1316 QLTLSDELLRLVYSSLDILKKSILSIP--AVOTLDDLEGVYVPGSPMODALFLOSQSK 1373

171 -----EV-F-----I-C--AM-----N-----P-----179

1374 DGAYEVDFTWRVATSLQNSQPAVDIGCLVEAKMDTVALHAALRTVILESSLPATGILHQV 1433

180 V--S--D--LL-----A-----S-----GS-----188

1434 VLRSHPDVIDLDRVDVTAITILDSYPPPTTEGIALIKRPPHRLICTTIEGSLVKFQ 1493

189 -----G--D-----S-----T--A-----193

1494 VNHLPDGMSTDKIIODLSKAYTCRHSNKLDPHSESKLHDGTGYNRPTKPLAEFIRYR 1553

194 --R--I--W-N-----L-----SE-----200

1554 DPQRKQDSINYKWNALRGATTCSPPLPDQITSEKAMPQSWASVPILCVDSKELSKYL 1613

201 -N-----ST-----S--G--S--T--Q-----LV-----210

1614 ANLGITMTMFQTVMAIVLRIYSQGSVFGYLTSGRDAPVDGIDSAVGNFIAMLVCFED 1673

211 L-----R-----H--C-----IRE--G-----218

1674 FDDGVHTVADMARKIHNASANSISHQACSLAEIQDALGLSTSTPLFNFTATYLPKRPTN 1733

219 --GO-----D--V--P--SNK-----D-----V--TS-----230

1734 VKAGEPEHLCELSMSDPTFEDTLTFVEPTQESNEVSAHLDFKLSYISQAVATSIAT 1793

231 -----L-----D-----WN-----234

1794 VAHILSELVHDVPRALNTLPVSEHDTAIRGNWDLHLPPTTECHETFSRKVHEHPQRE 1853

235 -----S--E-----GT-----238

1854 AICSDGSLTYAELSDLSQRLSHLVSLGKIVGTIKIPICEKSMWTIVTILAVVQAGGVF 1913

239 -----LL-----240

1914 VLLEPCHPSRLSGIHKVQAEALLCSPATSRMGALQNLSTQMGTEFKIVELEPEFIRSL 1973

241 -----ATG-----243

Db 1974 PLPPKPNHQMVGNDLLYVFTSGTGVKGAATHQAYATGIYEHAVACGMTSLGAPP 2033
QY 244 ---SY-D-G-F-A---R-----I-----W-T----- 253
Db 2034 RSLQFASYSFASIGDIFTTLAVGGCLCPREEDRNPAITTFINRYGVWAGITPSLAL 2093
QY 254 ---KD-----G-----NL-----A-----ST-- 261
Db 2094 HLPDPAVPTLKALCVAGEPLSMVVTWMSKRLNLINMYGTEATVACIANQVCTTTTYS 2153
QY 262 ---LG-----Q-H-----KG-----P-----IPALK--W- 274
Db 2154 DIGRGYRATVWVQPDHNSLVPFAGVAGELIIEGILCRGLYNDPERTAEVF-IRSPSML 2212
QY 275 ---N-----K-----K-----K-----K----- 276
Db 2213 HDLRPNSTLYKTGDVRYSDGKIIPIGRKDTQVKNQGFELGEVEHALQLQDPSDGP 2272
QY 277 ---K-----K-----GN--FI-----LS----- 283
Db 2273 IIVDLLKRTQSGRPDLIIAFLFVGRANTGTGNSDEIFATSTSSLSEFSTVIKKLQDAOR 2332
QY 284 ---T-TI---IW-----AG-VD-----K- 288
Db 2333 AMEVLPFMVPOAYIPIEGGIPLTAAGKIDRRMLKLCBPFNENDLISFTSKALSTSVKD 2392
QY 289 ---T-TI---IW-----AG-VD-----K- 293
Db 2393 AETTDIVERLARIEKVLGVKGVGRESDFPFSSGGNSMAIALRAEAQRSGFTLFVADIF 2452
QY 294 ---D-A---H-----G-----EA----- 296
Db 2453 TNPRLADMAKLFSGQSVSPSSSTLRKVPISLQKRSSGLOTAAPVNSGSPVRRCKEN 2512
QY 297 ---T-----T-----G-----EA----- 300
Db 2513 IIDCPVAFYEBGSPDTOLKEASRICGISRSIEDVFPCTPMOEALVALSLIPGAQASYA 2572
QY 301 ---K-----K-----K-----K----- 301
Db 2573 LHAAPFLRGLDRNRRFSAWESTVKAQPIRLSRHIIISGNSGVVVTSATDTSIPOLDVSGL 2632
QY 302 ---QO---F---P-----F-----H-----S---A--- 309
Db 2633 DTELEQOLQVGFAPGAPFLPRLAFVYSKADDCDFVISAHHAIYDGNLNLINQSVLAVT 2692
QY 310 ---P---A---LD-----V-D----- 315
Db 2693 NGELPPPGSPFKHFAARNLNLVQSKLDSDFWRKLLVKPQDESFRFPDVPVGHKPATRCTT 2752
QY 316 ---W---Q-SN--N-T---F-----F----- 322
Db 2753 NFHPFPMOSKIGTTANTCINAAMAITLAQYSSNKTVNFGLWGRDFPMIDIEHMTGPT 2812
QY 323 ---A---S-----S-----C-ST--- 327
Db 2813 IIVTPQVNVIPSSVAEFLODLQKSLAVVLPQHGLHRIQALGPIAQAOCDFSLVV 2872
QY 328 ---D-----M-----CI----- 331
Db 2873 NHGSSISWSELEAADIVPPLRSSDLAYPMVVEVENASSDITLIRVHSDPDCEVOLLE 2932
QY 332 ---H---VC-----K-----K----- 335
Db 2933 RLMEQFHNLTCLRAASFPDGKRIAEMLDDTATHTLTFWSNRVKSOPDVAIAVHK 2992
QY 336 ---L-----CO-----DR-----P-I----- 342
Db 2993 LLEETAQSPAESAIVAHDGQLSYMQMDRCADVLARQIRKTNMISAQSPFVCIHLRSAT 3052
QY 343 ---K---TF-----K----- 345
Db 3053 AVVSMVLAVLKAGGAFMPVDISQPRSLQNLIIEBSGAKLVLTLPESANALATLSGLTKVIP 3112

QY 346 ---Q-----G-----H-----T- 349
Db 3113 VLSLSELVQOITDNTTKKDEYCKSGDTPSSPAVLLYTSGTSGKPKGVVMEHRAWLSLFTC 3172
QY 350 ---N-----E-VN----- 353
Db 3173 HAEYMGFNSCTRILQPSLSLMDLSILEIWAIVLYAGGCLFIPSKDKERVNNLQDTRINDIN 3232
QY 354 ---A-I---K-----W----- 357
Db 3233 TVFLTPSIKLNPKDLPNISFAGFGEPMTRSLIDAWTLPGRRLVNSVGTPEACVLVTA 3292
QY 358 ---D-PT---G-N-----L-----A-S---C----- 367
Db 3293 REISPTAPHDKSSNIGHALGANIWWVEPQRTALVPIGAVGELCIEAPSLARCILANPER 3352
QY 368 ---SDD---TL-----K----- 367
Db 3353 TEYSFPSTVLNDNWQTKGTRVYRTGDLVRYASDGLDFLGRKDGOIKLRGORIELGEIH 3412
QY 368 ---SDD---TL-----K----- 371
Db 3413 HIRRLMSDDPRFHEASVOLYNPATDPRDATVDVQMRPEYLAGLLVLDVLSMRSDSMH 3472
QY 372 ---S-----M-----K----- 373
Db 3473 VLNIANTSNIQTLVTELKKSRLGVLPHYMVPLHFVAVSRLPTGSSGKLDHAFVACLRE 3532
QY 374 ---KI-----W----- 376
Db 3533 LTAPLDGNPFKEQVLTTNESVLRQWGTVLAMDPHSIQGRDDFPISLGSSISAMRLVGL 3592
QY 377 ---S-----M-----K----- 379
Db 3593 ARSSGHKLQHEDI FMCPRILADMAGQISFVQEAASVSPTTPTKFDLLDCEVEVIDHIL 3652
QY 380 ---Q-D-N---C----- 383
Db 3653 PQLDNNKELLIEDVYPTCPLQESLMAATARHGEAYTMIQSIITVLAQAQLKXAMDVVFRD 3712
QY 384 ---V-H---D-----L-----Q- 388
Db 3713 FEVLRTRIALGPSQOALQVWVKHELSWESFPISQSKDFHYSRSLGKPKLARLAVITQA 3772
QY 389 ---Q-H---N-----KE-----I-- 394
Db 3773 LDTKQPIHSGTREARTKNSQDVTVMVVGAHHSIYDAHLVLSMIWRLRYREFIGSQADGILE 3832
QY 395 ---Y----- 395
Db 3833 AETSREGVVPKSVYVEKLLRGKDNDESLLFWKEKLRGVSSSQFPASPASWPRVLEHQPSAT 3892
QY 396 ---T-I---K-----W-----S-P---T-G--- 403
Db 3893 QTLITKVSLPTSSRKKGATVATVAYAAWALTIAHYTADPDVVFAGTSLGRETMAGSISH 3952
QY 404 P---G-T-----N-----N----- 408
Db 3953 PESIAGPTIITVPLRIIIDFQTVWSDFLSTLQKDIVRAAYFGQMMGLNSIAHIDNDCRDA 4012
QY 409 ---P---N---AN-L-M-L---A-----SAS----- 419
Db 4013 CGFKSIIVQVQDEGENHVGRAANPFQMSLESIGHPAPLVEVESESTVDLIRMAYDP 4072
QY 420 ---F-D---S-----T----- 423
Db 4073 VLVPEKLAHFISDTFTTTTMSNLNLSAANPAKAKVESIPALSEAHLAELDVTCTPEMILGKAKDE 4132
QY 424 -VR---L-----WD----- 428
Db 4133 KIRTESHOCLODLVCCRRAQSPNSQAIDSDGSI SYHELDGLSLIAEHLSQLGVPEAP 4192

```
QY 429 -----V-D----- 430
Db 4193 VCLLFKSKWAVVAMIGIIKAGCGFPLDPSYPHERLEHIISETGSSVIVTSAAYSKLCL 4252
QY 431 -----RGI-C-----I-----H- 436
Db 4253 SLSVRGIVCDGVSFSTKPLPSTADSPSPFSVRPNQAAVILFTSGTGKPGVMEHHS 4312
QY 437 ----- 436
Db 4313 VCSALIALGRMGLGQSRVLFQNSYWFVMDLLIDIFGTLVYGCLCIPKEBQMSNLSGW 4372
QY 437 -----T-LT-----K-H----- 441
Db 4373 VQFKVNTMLLSTSVSRMLQPADTPSLETCLTGEAVLQSDVDVRMAPKHLIAGYPTET 4432
QY 442 -----Q-----E-----P-V-----Y----- 446
Db 4433 CIMSUGELTPSSPANLIGKPVSCQAWINPLAKETELAPYGATGELYIQGPTVARGYLHD 4492
QY 447 -----S-----V-----A-----F-----S----- 451
Db 4493 DVLTSKAFIVDPQWLTYGTNENQWSRRAYKTGDLVFWGQSNLYTVRRKDSQVKIRGQ 4552
QY 452 -----PD-----G-R-----Y- 456
Db 4553 RVELAEIEEVIROHIPDVTVCVDLLSSDDQNTRIILGAVLIGDRALGDPEDLEVIGYM 4612
QY 457 -----L-AS-----GS----- 461
Db 4613 DLLKSHIIPALEASLPHHMIPEAVVPVQLPTLIGSGKLDKRTVRRVAGPLAPSLPOASAR 4672
QY 462 ----- 461
Db 4673 HPNQPTVTHQKLLRLQWLKILPQDESAYVKQDNFLIGGDSIAAIAKLVALLRQHGISL 4732
QY 462 -----F-----D-----K- 464
Db 4733 AVAEIFTRPLEAMSSLDIHNFFVSHAGILSDVTRTSGVMRQTNLIAGRHMAVEKS 4792
QY 465 -----C-----VH-----I-----W-N-- 470
Db 4793 RECDNSTLPCTEYQOMELAGTEAFTGAHSAQIFRLPEKIDLRLOQAADFHCADWYPNLR 4852
QY 471 TQ-----V--CLHY--LN-----G--QV----- 482
Db 4853 TQIHKDADTGRLLHDISPVGKVPWSC-HYSDDLNTVLSHDKPPPGDGLPLHRVTIMRH 4911
QY 483 -----L--LN-----LG----- 487
Db 4912 RDPTEMLVWTLNHAAYDAWSLRMLLEHYTEAVANPDYEPSYSLGWTAFVLHTENTKEAS 4971
QY 488 RS-----I----- 490
Db 4972 RSFWSSYLDVKPARLMFNVLNPNRQDRLYEARINIPKRVLSQATAATVLLAGLTLV 5031
QY 491 -----C-----LYT----- 494
Db 5032 ARVCDTRDVLHLLTGRITLPLAGIENCPGPTITKVPRLPLMDQDLVLTLELDSVAKKIT 5091
QY 495 -----L--PH-H--L-----V-----VI-P----- 503
Db 5092 AELMRVPHPHISGLSATSAREFTPOAEGTTSSGKPHAGSVLGRPLDLVIHPKGGDLILGK 5151
QY 504 -----LVA-----L-----TEL- 510
Db 5152 HGLGLQNEGFRFLVAPPSGGLSMECALVDDDDKRSDTISVDVSVLWDQRAATQBDVIELV 5211
QY 511 --L--VL-K 514
Db 5212 HSLQGIETK 5220
```

```
RESULT 10
S58870
reelin precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Dec-1996 #sequence revision 06-Dec-1996 #text_change 21-Jul-2003
C:Accession: S58870; S71844; I49297
R:D'Arcangelo, G.; Miao, G.G.; Chen, S.C.; Soares, H.D.; Morgan, J.I.; Curran, T.
Nature 374, 719-723, 1995
A:Title: A protein related to extracellular matrix proteins deleted in the mouse mutant
A:Reference number: I49297; MUID:95231649; PMID:7715726
A:Accession: S58870
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3461 <DAR>
A:Cross-references: EMBL:U24703; NID:G902486; PID:G902487
R:D'Arcangelo, G.
submitted to the EMBL Data Library, April 1995
A:Reference number: S71844
A:Accession: S71844
A:Molecule type: mRNA
A:Residues: 1-215, 'T', 217-1905, 'S', 1907-3355, 'V', 3357-3391, 'N', 3393-3461 <DA2>
A:Cross-references: EMBL:U24703; NID:G902486; PID:G902487
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-3461/Product: reelin #status predicted <MAT>
F:1769-1795/Domain: EGF homology <EGF>

Query Match 70.4%; Score 2568.3; DB 2; Length 3461;
Best Local Similarity 11.5%; Pred. No. 7,7e-53;
Matches 395; Conservative 94; Mismatches 13; Indels 2947; Gaps 362;

QY 1 M-----SI-----S-----S-----D-----EV----- 8
Db 1 MERGCWAPRALVLAIVLLALLATLRARAATGYPRFPFFFLCTHHGELEGDGQGEVLISL 60
QY 9 ---N-----F--LV--YRY---LQ--ES-----GF---S-H 23
Db 61 HIAGNPTYVYVQGEYHVITISTTFDFGLLVTLG-YTSTSIQSSQSGSGSFAFGIMSDH 119
QY 24 ---S-A-----F-----T-----P-----G-I-K----- 31
Db 120 QFGNQPMCSVVAHVSHLPTNLSPFWIAPPACTGCVNPMATATHRGQVIFKDALAQQLC 179
QY 32 -----SH-----I-----S-Q-----SNI-----N-----GAL----- 43
Db 180 EQGAPTEATAYSHLAIEHSDSVILRDDFSYQQLLELNPINWECNCEMGEOGCTIMHGN 239
QY 44 -V--P--P-----A-AL--I-----SI-----IQ-- 54
Db 240 AVTFCEPYGPRELTTCLNTTASVLQFSGISGSCRFSYSDPSITVSYAKNNTADWILE 299
QY 55 -----KG--LQY-----V-BA-E-----VSIN----- 67
Db 300 KIRAPNSVSTVHILYLPPEAKGESVQFQWKQDSLVRGEVVEACWALDNLIV-INSARE 358
QY 68 ---ED-----GT-LF-----DG-----R----- 76
Db 359 VLELDNLDPDVTGNWLFPPGATVKHSCQSDGNSIYFHNESGEFNFATRDVLDLSTEDIQ 418
QY 77 -----P---I-----ES-LSLI-----DA----- 86
Db 419 EQWSEFEFSQPTGWDILGAVVGADCGTVESGSLVFLKDGKRLCTPYMDTGYGNLRFY 478
QY 87 -VM-----P-----D-----V-----V----- 94
Db 479 FVNGGICDQGVSHENDIILYAKIEGRKEHIALDTLTYSSVKVPSLVSVVINPELQTPATK 538
QY 95 ---RQ---QAY-R---D-----K-LAQ-----Q----- 106
Db 539 FCLRQKSHQGYNRNVWAVDFHVLPLVLPSTMSHMIQFSINLGCCTHQPQGNVSLEFSTNH 598
QY 107 -QA-----A-----AAA-----AA----- 114
Db 599 GRWSLLHTECLPEICAGPHLPSTVYSSENYSGWNRITPLPNAALTRDTRIRWQTGP 658
```


Db 2810 VFPTGKRIITYPLPESLTGNVPRFRFYKYSDVQWAINDFYLGCGCLDNCGGHGDCLK 2869
QY 437 -----T-----ITKH-----QE--P-----V-----445
Db 2870 EQICIDPGYGNPCYLT-HSLKTKFLKRFDSBIKPDLMWLEGGSTCTCGVLAENTAL 2928
QY 446 Y---S-V-----A-F-----S--P-----DG-----454
Db 2929 YFGSTVROAITQDLRGAFLQYWGRIGSENNMTSCHRPVCKBGLVDFDGTGGITW 2988
QY 455 -----R--Y--L--A-----SG-----460
Db 2989 TLLHMDFOKYISVRHYILLPGALTTNTRLRWQPFVISNLVSGVGACAVGTGQHS 3048
QY 461 -----S-----FD-----K-----C-----VHI-W-N-----T-----471
Db 3049 DWSNRNHPQSOLVDTFDEGSSHEENWSFYDPAVRTAGFCGNPFSFLYWPNNKKDKTHNAL 3108
QY 472 -----Q-----V-----C--LH-----Y-----L-----N--479
Db 3109 SSRELIQPGYMQFKIVVCEATSCDGLHVMLEYTKDARSOWOLVOTQCLPSSNSI 3168
QY 480 G-----Q-----VL--LN-----G-----R-----SI--490
Db 3169 GCSPPQFHEATIYNANSSSKRITIQLPDHVSSATQFRWIKGBETEKQSWAIDHVI 3228
QY 486 -----L-----G-----R-----SI--490
Db 3229 GEACPRLCSGHGYCTTGAVCIDCESFGDDCSVFSHELPSYIKDNFESARVTEANWETIQ 3288
QY 491 -----C-----LY-----T-----494
Db 3289 GGIVGGCOLAYAHGDSLYFNGCQIROAATKPLDLTRASKIMFVLQIGSPAQDSCNS 3348
QY 495 -L--PH-----HLV-----V-----IPL-----V-----505
Db 3349 DLSGPHADKAVLQYSVNNGITWHVIAHQHPKDFQCAQVRSVHVPLEARMKGVLLRWQ 3408
QY 506 -----AL--IB-LIV-----LK 514
Db 3409 PRHNGTGHQDQWALDHVEVVLVSTRKQNYM 3437

RESULT 11
T12117
polyprotein - fava bean dsRNA replicon
C:Species: Vicia faba (fava bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12117
R:Pfeiffer, P.
J. Gen. Virol. 79, 2349-2358, 1998
A:Title: Nucleotide sequence, genetic organization and expression strategy of the double
A:Reference number: Z17424; MUID:98451319; PMID:9780039
A:Accession: T12117
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5825 <PFE>
A:Cross-references: UNIPROT:O82731; EMBL:AJ000929; NID:g3184155; PIDN:CAA04392.1; PID:g3
A:Experimental source: virion; cultivar 447
C:Comment: This gene product may be cleaved into several proteins including helicase and
C:Genetic:
A:Genome: dsRNA replicon
C:Superfamily: fava bean dsRNA replicon polyprotein
Query Match 70.3%; Score 2566.1; DB 2; Length 5825;
Best Local Similarity 9.3%; Pred. No. 2.6e-52;
Matches 441; Conservative 57; Mismatches 13; Indels 4239; Gaps 379;
QY 1 M-----SIS-----S-----D-----E-----V-----N--9
Db 949 MATRPSITQAPVYKMSNTLLIDMNLIRMDPMWLITECLRRKITTATYVTPMELDONPD 1008

QY 10 -----F-----LV-----Y-----13
Db 1009 QTILDEGVQFRKVENQIYVITAWSDRPIRVTTENYNNWTKVDAFVIGENTYPIDTVREAN 1068
QY 14 -----R-----Y-----L-QE--18
Db 1069 GISLRIDLAASTPMSSEWVETVGDSEISVPTVYVYKPGKRKLSMGSEKKVHLREL 1128
QY 19 -----SG-----PS-----HS-AF-----26
Db 1129 YRMMIRNVSGTNSFKELVDYIGISNMKNYNDVMVVSFSDISDDTIRHTAAAFVSKMIT 1188
QY 27 -----T--F-----28
Db 1189 RANSMQLTERTVLESVNTARGFLASLLISOAETMLKITGLDETMLTTVNRWMSKSEWVRL 1248
QY 29 -----G-----IK-----31
Db 1249 KGMLMNLISLDIDNLWCYSREPVIHRLAQETLELEPTGTCPHKTPDFVIKECSMCECCG 1308
QY 32 -----S-----H-----33
Db 1309 VNPAGIAGFCSSCLPTDKHQCDHPCKHAHESEAGTRCSCCLPIAGEACPCGVNRQIE 1368
QY 34 --I-----S--Q-----O-----SNIN--40
Db 1369 SEILPENSESADAEEQNRVKNKPRPRDREKGNRNNRNDSSRRTDANHHISNVYHH 1428
QY 41 -----GALV--PP--AAL-----I-----S-----I 52
Db 1429 GHNKPORQGA-TQPPKRSALHPENDNDTPDSTPVLITADPTTNPGERQAPGDNI 1487
QY 53 --I-----Q-----KG-----L--Q-----Y--V 60
Db 1488 PHENDIPGSSSTTQSSPPDDTNYSGPEHNRIQNHRIIAQTNFEFGDLAYAGNDIAV 1547
QY 61 -----E-----A--EV--I--S-----I--N 67
Db 1548 KKNSHKLPNPEORLRSCEFQNMGDSIYATENLEVIKTYNTSSEGEGYCGYNALKILYPN 1607
QY 68 -----E--D-----69
Db 1608 LDLTLEMQEIVGSETQFSDWEIMRVAQAKQLNLI VVTERCALVNSVCSNBEFGVICHCR 1667
QY 70 --G-----TL-----F-----D--G-----R-----P--77
Db 1668 HRGVMLHWEAALAIQKGFADYHPTTNALTRDLDLDFAKSSGLNKNKVRSFVLGDRLL 1727
QY 78 -I-----ESLS-LI-D--A--VM-----P-----D-----90
Db 1728 KLQTEMHESLSKVVEDVNPAGFVKITKGSTHYVTVNDRNRLGYSPQYGTFTAQITDNOEL 1787
QY 91 -----VVQT-----R-----Q-----QA-----98
Db 1788 VELLILAYSAPARVIHTDFYDRPESVPADATQGIHDYKQLVRDIAEVNQACAGDLIKS 1847
QY 99 -Y--R-D-K-----L-AQ-----Q-----O--107
Db 1848 ELIQNRVDCKEYINNHLFPVKQTKGLKPGDLISGLKVLGIQSTSYVDVSHVIGVIRTNQL 1907
QY 108 -----A-A-----A-----A-A--A--A--114
Db 1908 WCTCVTEKCAKATKIKVDLYRNKVTGSLRALFGIFRWDFFRDIKTLEKATAVDIAGWG 1967
QY 115 -----A--AAAS-----QQG-----S-----123
Db 1968 KSTEIVKLVNQDCTVVAQTSAAVSNILEKLEQKKMKCKVMSIEKMTQVQNTPTLVLD 2027
QY 124 -A-----KN-----G-----E-----128
Db 2028 EASMITWETLSLITGQVENVYLGNTLQICVLDWYRTGSGRATKSIIEQAGIIRHYTT 2087
QY 129 ---N-----T-A--N-----G--BE-----135

Db	2088	HRIGNFLARELSLVTKELITNAKHETNCTKSWDAVRWAELETSTIAGSLEBPVILCFYNA	2147	Db	3167	YGRSYLMTQVYTYRYVNVINKNGLDLEAVCIAALTMKHIRNEQLEHTMGLMETFRYVD	3226
QY	136	---N---G-A---H---	140	QY	266	---K-G-P---I---F---	270
Db	2148	VRAVMYLVKGVCRDTHHKFOGLEADNVVLQWCPGTGPPGRITLDKHQCCLSAATRAKON	2207	Db	3227	DSITDIVKSCKNLEKVMSTANTMOEKLGPASFTGLLDEANTQORVVVFIDERTNQSR	3286
QY	141	---I---ANN---HT---DM---M---	149	QY	271	---A---LKW-N---KK---	277
Db	2208	LVWISVNEYSNVPLHKRMGATIGGSKHTQPDPTENNADKTLQVSHLLSKWMIYSKDKT	2267	Db	3287	VMSWDSNIKHSDEGSIKKWMASTADNLDMMNIGMIQEVYKVLCKREDIPKTKTKACPGF	3346
QY	150	---E---VD---G---	153	QY	278	---GN---	279
Db	2268	IERLECSRLQSOBEGNLVMLKNLTLHLDVYVSGNDQHSRLRQSGKLAKEFKFSMLSLFS	2327	Db	3347	KQVVDCAKAVAOQKEARKTDPGARVLVLEPIPLIGNVDITRAATOGTMVTKNENVIRSE	3406
QY	154	---D-V---EI---	157	QY	280	---F---L-S---AGVD-K---T-TII---	292
Db	2328	KEPAPDHVSSSTLGLKAEIVDQANANKDRINSQIMNIYTVQKNLAISNGSGIAAATPVVL	2387	Db	3407	ITFRARQSRRAQTSRLRSYLDNWNEMVKAYSALEAG-DLKVMVRELSRIALIKSTSTP	3465
QY	158	---P---N-K---AVV---LR---	166	QY	293	---D---A-H---T---GEA---KQ---Q---F---	304
Db	2388	TAMLAQKPGITKAECSDPEPTETLTANILMLGVNVLKIQLDFENDKVILGSGGLRSFVAD	2447	Db	3466	TWHLKEDLEAMLDVPNLTRAHLEITKLIGDAVVSWEQSKAEKKQNEVDFIKSDLGLST	3525
QY	167	---GH---E-S---E---V-F-IC---	175	QY	305	---P---	305
Db	2448	SSEIKOLYNQKFCNCGHYVLKYRNFDTETISVKKMNFERVLWTVNFGVCLSGSLHSV	2507	Db	3526	LNDHDLNIVLSLLAVDRPSKVKKDLLTNTLGSQMSYNEAVMLRRKIROQWCESASNGOL	3585
QY	176	---AW---N---P---V---SD---LLA 185		QY	306	---PH---S---A---	309
Db	2508	TIENELYKMQSYGCSLGGMWIFPHNKPIIVGDPSYNTGNSRTVFNFSKQSDKCPILA	2567	Db	3586	VDHHDHANSVSTTEQNOQKLNILSLHEEBAGICDINLSDASPLKIRASSYVLPRAITTCVQ	3645
QY	186	---SGS---GDS-TA---RI---W---	196	QY	310	---P---ALDVD---W---OS---	318
Db	2568	YLRLNVDSGTGVKLPYDPVGRGDFMEGDNVMAALVIERVLKGMWMAAKLKNMLTN	2627	Db	3646	TPGMVELTSPKVEIIEAMDNDBCVWRCIEKYVTQNIIEPHFRITNGRLAIMQOSKMLTES	3705
QY	197	---N-L-S---E---NS---T---SG-S---TOLV- 210		QY	319	---N---T---F---AS---C---STD---	328
Db	2628	TNODLCNFYDNNSPFLFDSLIKARLESNGFTNFNTLVNFBVGGTRLEFNGFTNFNTLVN	2687	Db	3706	QAIIVCQLLGLACCLIQNGETGTVYNFAPNKPQVQLMRLSQDASDHCVLNLNLGADGVK	3765
QY	211	---L-BH-C---I---	215	QY	329	---M---C---I---HV---C---	334
Db	2688	VNSMDLRCPILCTDKRYLAALMDQDNLIWTVQISGKPIHTPAGVTRVSNKAFKTMAYR	2747	Db	3766	RUSPENIAKENMEQLEHVCVDNKPVAVGENPYACVSHVHDNEDLARLTSCSQPYEKI	3825
QY	216	---R-E---G---Q---D---	221	QY	335	---	335
Db	2748	QMLNRLFKLTWQLSRDKPMGAPNPRGPIKKFMAQHEADNDKVAVAARKYKQTOAAMM	2807	Db	3826	PVLDRGFEMLPVNNVMARMTRKTSSATLLKKVTPITTTAGSGLATRYNYKATPMVTKPG	3885
QY	222	---V---P-SNK---D-V---	229	QY	336	---L---G-Q---	338
Db	2808	KLKSRMYLTAWVNGSEYASIMQOHNVITTTSTNYPENMLGLVDLYVAQAIHNTGYT	2867	Db	3886	DVLVLRDHCEKKVMCTHTTIVDTPGGVQIILEITDDIGDGAVCVLDLGLHCLTPADRTPR	3945
QY	230	---L-DW---N---SE---	236	QY	339	---DRP---IKT-F---QG---H---T 349	
Db	2868	SALYITNACTAVLCHGHWDFSCPPDDGWSRFLSSTHTDTITNLADIKSIILDSIAKLD	2927	Db	3946	RDRPNTSNEVWSLNIETORYLEATKGLRSHVQGGEGKLHIYHFDNREHHMYDEHYDA	4005
QY	237	---	238	QY	350	---N-E---VN---A---IK-W---D-P---	359
Db	2928	KQELSEETAVLKEKERSERIHNQILRTEDSWYSESYQNVKRGIVIVSVNSCGISTES	2987	Db	4006	MLDENIVIELDPTVQVPTGVNIVNLICIRRMFLAMLYGIPTWVAYCDNPSAAQLAWFL	4065
QY	239	---L---A---TGS---	245	QY	360	---	359
Db	2988	IEKIMETTGAEFLMVCVPTLTPSDNTTAHRVGTGNSNNKITVPGARNTLTNPETVYCM	3047	Db	4066	KYYKMGHKVFNNNVTEDLTNTRAMEIINALFSTYQHTTTTHTTRVHLRKNKAETILSKEK	4125
QY	246	---DGF---A---R---	250	QY	360	---T---G---N---CS---DD---	364
Db	3048	KSGKRATVOSQTKD-FDMISLHTTTVAMIASVQLIKVLTITRFRNEHAMPKALPIPFSPFS	3106	Db	4126	VSEELFNTTIDISVFTAVLMGLENEVLKSNQGTCTVFDLIPETPEPPLOQDMLESIGLV	4185
QY	251	---I-W---T-KDG---	259	QY	365	---A---S---CS---	370
Db	3107	HOSDEKVVQVPRFTCNFGDGSNSLIITKRRKLVSKALLAKMNERALRHDSNLKDLQA	3166	Db	4186	DIKQCAVHVPTGTKVLRVVSLSHEIACSTKYKGVDPKDDDEMAGASNLTSQVPSADK	4245
QY	260	---S---T-L---G---Q-H---	265	QY	371	---M---T---	375
				Db	4246	ADPGSSQLSSCBENNSQTGTQPVACVHI8PCDHAANASNDVLKEITQPEQNNHDIISKI	4305

```
QY 376 -----W-----S-----M----- 378
Db 4306 LATAQAKSNPEKSKGATSADLOAIQSLCDSPDGMGSQVADQAKKLEVMNFSDEVT 4365
QY 379 -----K-Q-----D----- 381
Db 4366 LSGIHRQRYVNSQGYLQKVDISERWAIYKIGIEHSGSDVPTNPNPGWWLAFPPMKRLG 4425
QY 382 -----NC-----V-H-----D-L-----Q----- 388
Db 4426 VYKEHEIKVMYAGCAVRLNRIKREPKSVFTINVKPEVMDLILTYTSANNKQYPTN 4485
QY 389 -Q-----H-----N-----K-E-I-----Y----- 395
Db 4486 LQRLADKYLHGRVSHVYVYEGYNRDFLOTVCBELTKRMRVIVVTDPDVYNQLDADIQV 4545
QY 396 -----TIK-----W-----SP-----TG-----P----- 404
Db 4546 REKIITIKVTSQPEPYRAIWRFMILFANIDPCELYILNGREMAFNDMELVDVRMEESF 4605
QY 405 -----G-----T-N-----N----- 408
Db 4606 EDFDIIVSALHCOAFNCCGVAKYNHAKFTDYNFTNDIYEAAGNSYGGDELYMSRYKVXW 4665
QY 409 -----P-NA-----NLML-----A-----SASF----- 420
Db 4666 WLYYILPGGAWYGIEGKWLNYALDTTICLKEGERTRIDHVKGLFCSANFHVADPTKG 4725
QY 421 -D-STV-----R-----L----- 426
Db 4726 IYDSTVDSVGDGPIQYAMQINGHSVTENPVHSTGSGTGNPRISNDGEPWQLNPN 4785
QY 427 -W-D-----V-D-----R-G-I-----C----- 434
Db 4786 NWLLDPLCKSKMKIQKFLDSSVKIKESQSLPIQLGPKLPVIVFVNAQCEAIGVSM 4845
QY 435 -----I-H-TL-----T----- 439
Db 4846 NNRGTLYVEVNIQAOKLAARNEPLOHQTLIKSHLTLMSTTGRHAGWKPEYTHLMKADLQ 4905
QY 440 -----K-----H-Q-----E-----P-----V----- 445
Db 4906 NCQYVEGSATSLWKRBTQHAQVYTTTEERDVFRIPQALDHSVGLTKLVKVTGNE 4965
QY 446 --YS-----V-----A-----F-----S-----P----- 452
Db 4966 LGLSSSGVLARVLKLNDRLANHWHLETOHLQKLEFPDRRIISHSSESGMTFKPRIGOT 5025
QY 453 -D-----D-----G-----R----- 455
Db 5026 LYLADVLTAAGIECVNHTTTSREWKEALKNPLTKVGIMPLTNSVLVSGKLCVQPTIPR 5085
QY 456 -Y-----L-----L-----AS----- 459
Db 5086 CYAMLCQGVKLIKFKPETRGGTGPSNVANVGSNPTTAGPGDTPPLITSWNKWSGAS 5145
QY 460 -----G-S-----FD-----K----- 464
Db 5146 NQSPFGVSAYQDKLREKKIETPDRLRVLEMTLFDFAVALMSHGLTLDRKLDNGGNWH 5205
QY 465 C-----V-----H-----I-----W-----N----- 470
Db 5206 CATDATMKSPFIEIPEYVGGDLLPSLNTSALRHNETNARVIDLWDDTDLRDLWLTLYAPKNP 5265
QY 471 -T-----Q-V-----C----- 474
Db 5266 MKLTSRVSPOGKINLKLTLNRCQTRPVPQVMGGENAVTCRLGSLVPLRREPNNVT 5325
QY 475 --LH-----Y----- 477
Db 5326 HELHKFTAYYDGRWVRLKDFKANTITISDADVKTWLSRRSDWKALATSTIKMLTGLP 5385
```

```
QY 478 -----L-N----- 479
Db 5386 SNPMNAVNVHTESLLKANPIWYRQTOGRIIVWQPKELCALMSPAFIAIKRLEKVL 5445
QY 480 -----G-----QV-----L-----L----- 484
Db 5446 DEIVYTDGLTPDMLSARARTIQDYVVFEDDLVIQDRQTDQELIDLEFQVMLDLGLDINLA 5505
QY 485 NL-----G-----R-----SI-----C----- 491
Db 5506 NLWLRLVHNKWRPKFGQHSWGQLDAMBLTGOATTALGNATNLVHSSFVIEHROAIKLMFV 5565
QY 492 -----L-----Y-----T-----L-----P----- 496
Db 5566 LGDDNITFMSAPNLTKYKRLMSERYNMRSKQVSRNVCTFCSLLCYRNSFGHCEVGPDP 5625
QY 497 -----H-----H-L-VV-I-P-----L-V-----A-L-----IE-L----- 510
Db 5626 VRLRHRFEVTGGNHVELPVAVKMHPHNVIRIGKTFNGRYLHVMEENEAEAEPEVWLRNH 5685
QY 511 -----LVLK 514
Db 5686 ANRSYILVLK 5695
```

RESULT 12

S02392

alpha-2-macroglobulin receptor precursor - human

N;Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C;Species: Homo sapiens (man)

C;Date: 14-Aug-1998 #sequence revision 14-Aug-1998 #text_change 09-Jul-2004

C;Accession: S02392; S30027; I37998; A39210; S12538

R;Herz, J.; Hamann, U.; Rogne, S.; Myklebost, O.; Gausepohl, H.; Stanley, K.K.

EMBO J. 7, 4119-4127, 1988

A;Title: Surface location and high affinity for calcium of a 500-kd liver membrane protein

A;Reference number: S02392; MUID:89210795; PMID:3266596

A;Accession: S02392

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-454 <HR>

A;Cross-references: UNIPROT:Q07954; EMBL:X13916; NID:G34338; PIDN:CAA32112.1; PID:G34339

R;Kristensen, T.

submitted to the EMBL Data Library, October 1990

A;Reference number: S30027

A;Accession: S30027

A;Molecule type: mRNA

A;Residues: 3275-3864 <KRI>

A;Cross-references: EMBL:X55077

R;Herz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.

EMBO J. 9, 1769-1776, 1990

A;Title: Proteolytic processing of the 600 kd low density lipoprotein receptor-related p

A;Reference number: S12538; MUID:90269210; PMID:2112085

A;Contents: annotation; site of proteolytic cleavage

R;Kutt, H.; Herz, J.; Stanley, K.K.

Biochim. Biophys. Acta 1009, 229-236, 1989

A;Title: Structure of the low-density lipoprotein receptor-related protein (LRP) promoter

A;Reference number: I37998; MUID:90089395; PMID:2597675

A;Accession: I37998

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-11 <RES>

R;Cross-references: EMBL:X15424; NID:G34408; PIDN:CAA33464.1; PID:G34409

R;Strickland, D.K.; Ashcom, J.D.; Williams, S.; Burgess, W.H.; Migliorini, M.; Argraves,

J. Biol. Chem. 265, 17401-17404, 1990

A;Title: Sequence identity between the alpha2-macroglobulin receptor and low density lip

A;Reference number: A39210; MUID:91009181; PMID:1698775

A;Accession: A39210

A;Status: preliminary

A;Molecule type: protein

A;Residues: 150-166;234-238, 'X', 240-245, 'X', 247-252; 'G', 686-695;902-916;1096-1109; 'S', 17

C;Genetics:

A;Gene: GDB:LRP1; APR; LRP; A2MR

A;Cross-references: GDB:119694; OMIM:107770

A:Map position: 12q13.1-12q13.3
 C:Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated proteins (see PIR:A39875).
 C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding site; signal sequence #status predicted <515K>
 C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glycoprotein
 F:1-13/Domain: signal sequence #status predicted <515K>
 F:20-3943/Product: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>
 F:27-64/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:72-108/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:115-148/Domain: LDL receptor ligand-binding repeat homology <EG1>
 F:154-188/Domain: EGF homology <EG2>
 F:198-239/Domain: LDL receptor YWTD-containing repeat homology <YW01>
 F:240-281/Domain: LDL receptor YWTD-containing repeat homology <YW02>
 F:292-334/Domain: LDL receptor YWTD-containing repeat homology <YW03>
 F:335-378/Domain: LDL receptor YWTD-containing repeat homology <YW04>
 F:379-420/Domain: LDL receptor YWTD-containing repeat homology <YW05>
 F:421-468/Domain: LDL receptor YWTD-containing repeat homology <YW06>
 F:478-519/Domain: EGF homology <EG3>
 F:571-613/Domain: LDL receptor YWTD-containing repeat homology <YW07>
 F:614-659/Domain: LDL receptor YWTD-containing repeat homology <YW08>
 F:660-710/Domain: LDL receptor YWTD-containing repeat homology <YW09>
 F:711-752/Domain: LDL receptor YWTD-containing repeat homology <YW10>
 F:753-799/Domain: LDL receptor YWTD-containing repeat homology <YW11>
 F:807-842/Domain: EGF homology <EG4>
 F:854-890/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:895-931/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:936-971/Domain: LDL receptor ligand-binding repeat homology <LDL5>
 F:976-1011/Domain: LDL receptor ligand-binding repeat homology <LDL6>
 F:1015-1051/Domain: LDL receptor ligand-binding repeat homology <LDL7>
 F:1062-1097/Domain: LDL receptor ligand-binding repeat homology <LDL8>
 F:1104-1140/Domain: LDL receptor ligand-binding repeat homology <LDL9>
 F:1145-1182/Domain: LDL receptor ligand-binding repeat homology <LDLA>
 F:1185-1221/Domain: EGF homology <EG5>
 F:1227-1261/Domain: EGF homology <EG6>
 F:1269-1308/Domain: LDL receptor YWTD-containing repeat homology <YW12>
 F:1309-1355/Domain: LDL receptor YWTD-containing repeat homology <YW13>
 F:1356-1398/Domain: LDL receptor YWTD-containing repeat homology <YW14>
 F:1399-1445/Domain: LDL receptor YWTD-containing repeat homology <YW15>
 F:1446-1488/Domain: LDL receptor YWTD-containing repeat homology <YW16>
 F:1489-1531/Domain: LDL receptor YWTD-containing repeat homology <YW17>
 F:1540-1578/Domain: EGF homology <EG7>
 F:1583-1626/Domain: LDL receptor YWTD-containing repeat homology <YW18>
 F:1627-1669/Domain: LDL receptor YWTD-containing repeat homology <YW19>
 F:1670-1713/Domain: LDL receptor YWTD-containing repeat homology <YW20>
 F:1714-1753/Domain: LDL receptor YWTD-containing repeat homology <YW21>
 F:1754-1796/Domain: LDL receptor YWTD-containing repeat homology <YW22>
 F:1797-1846/Domain: LDL receptor YWTD-containing repeat homology <YW23>
 F:1850-1886/Domain: EGF homology <EG8>
 F:1934-1976/Domain: LDL receptor YWTD-containing repeat homology <YW24>
 F:1977-2019/Domain: LDL receptor YWTD-containing repeat homology <YW25>
 F:2020-2063/Domain: LDL receptor YWTD-containing repeat homology <YW26>
 F:2064-2105/Domain: LDL receptor YWTD-containing repeat homology <YW27>
 F:2106-2151/Domain: LDL receptor YWTD-containing repeat homology <YW28>
 F:2159-2194/Domain: EGF homology <EG9>
 F:2199-2241/Domain: LDL receptor YWTD-containing repeat homology <YW29>
 F:2253-2294/Domain: LDL receptor YWTD-containing repeat homology <YW30>
 F:2344-2388/Domain: LDL receptor YWTD-containing repeat homology <YW31>
 F:2389-2429/Domain: LDL receptor YWTD-containing repeat homology <YW32>
 F:2430-2473/Domain: LDL receptor YWTD-containing repeat homology <YW33>
 F:2482-2517/Domain: EGF homology <EG10>
 F:2524-2561/Domain: LDL receptor ligand-binding repeat homology <LDLB>
 F:2566-2600/Domain: LDL receptor ligand-binding repeat homology <LDLC>
 F:2605-2639/Domain: LDL receptor ligand-binding repeat homology <LDLD>
 F:2652-2688/Domain: LDL receptor ligand-binding repeat homology <LDLE>
 F:2696-2730/Domain: LDL receptor ligand-binding repeat homology <LDLF>
 F:2734-2769/Domain: LDL receptor ligand-binding repeat homology <LDLG>
 F:2774-2812/Domain: LDL receptor ligand-binding repeat homology <LDLH>
 F:2816-2853/Domain: LDL receptor ligand-binding repeat homology <LDLI>
 F:2858-2897/Domain: LDL receptor ligand-binding repeat homology <LDLJ>
 F:2904-2939/Domain: LDL receptor ligand-binding repeat homology <LDLK>
 F:2944-2980/Domain: EGF homology <EG11>
 F:2986-3021/Domain: EGF homology <EG12>
 F:3029-3068/Domain: LDL receptor YWTD-containing repeat homology <YW34>

F:3069-3113/Domain: LDL receptor YWTD-containing repeat homology <YW35>
 F:3114-3156/Domain: LDL receptor YWTD-containing repeat homology <YW36>
 F:3157-3200/Domain: LDL receptor YWTD-containing repeat homology <YW37>
 F:3201-3241/Domain: LDL receptor YWTD-containing repeat homology <YW38>
 F:3242-3284/Domain: LDL receptor YWTD-containing repeat homology <YW39>
 F:3294-3330/Domain: EGF homology <EG13>
 F:3334-3369/Domain: LDL receptor ligand-binding repeat homology <LDLL>
 F:3374-3408/Domain: LDL receptor ligand-binding repeat homology <LDLM>
 F:3413-3448/Domain: LDL receptor ligand-binding repeat homology <LDLN>
 F:3453-3489/Domain: LDL receptor ligand-binding repeat homology <LDLO>
 F:3494-3531/Domain: LDL receptor ligand-binding repeat homology <LDLP>
 F:3536-3570/Domain: LDL receptor ligand-binding repeat homology <LDLQ>
 F:3575-3609/Domain: LDL receptor ligand-binding repeat homology <LDLR>
 F:3613-3647/Domain: LDL receptor ligand-binding repeat homology <LDLS>
 F:3654-3690/Domain: LDL receptor ligand-binding repeat homology <LDLT>
 F:3695-3731/Domain: LDL receptor ligand-binding repeat homology <LDLU>
 F:3741-3776/Domain: LDL receptor ligand-binding repeat homology <LDLV>
 F:3785-3822/Domain: EGF homology <EG14>
 F:3828-3860/Domain: EGF homology <EG15>
 F:3868-3911/Domain: LDL receptor YWTD-containing repeat homology <YW40>
 F:3912-3969/Domain: LDL receptor YWTD-containing repeat homology <YW41>
 F:3944-4004/Product: alpha-2-macroglobulin receptor 85K chain #status predicted <85K>
 F:3944-4004/Domain: 85K chain extracellular #status predicted <85K>
 F:3970-4012/Domain: LDL receptor YWTD-containing repeat homology <YW42>
 F:4013-4056/Domain: LDL receptor YWTD-containing repeat homology <YW43>
 F:4057-4099/Domain: LDL receptor YWTD-containing repeat homology <YW44>
 F:4100-4142/Domain: LDL receptor YWTD-containing repeat homology <YW45>
 F:4151-4182/Domain: EGF homology <EG16>
 F:4200-4231/Domain: EGF homology <EG17>
 F:4236-4267/Domain: EGF homology <EG18>
 F:4272-4303/Domain: EGF homology <EG19>
 F:4308-4339/Domain: EGF homology <EG20>
 F:4344-4374/Domain: EGF homology <EG21>
 F:4377-4408/Domain: EGF homology <EG22>
 F:4421-4444/Domain: transmembrane #status predicted <TM>
 F:4445-4454/Domain: intracellular #status predicted <INT>
 F:166,2998/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
 F:2958/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
 F:4075,4125,4276/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 70.2%; Score 2562.6; DB 1; Length 4544;
 Best Local Similarity 10.4%; Pred. No. 1.9e-52;
 Matches 414; Conservative 79; Mismatches 16; Indels 3474; Gaps 366;

QY	1	MS-----I-----SSDE-----V-----N--F-L-----11
DB	88	MSRLNGVQDCMDGSDGPHCRELQNC SRLCGQHHCVP TLDGPTCYCNSSFQADGKT 147
QY	12	-----VY-----R-----Y-LQ-----E-----18
DB	148	CKDFDECSVYGTCSQLCTNTDGSFICGCVGYLLQPDNRSCAKQNEPVD RPPVLLIANSQ 207
QY	19	-----SG-----F5-----H--SA-----P 26
DB	208	NILATYLSGAQVSTITPTSTROT TAMD FSYANETCVH VHGDSAAQTOLKCARMPGLKGF 267
QY	27	-----TFGK-S--HLSQ-----S-N-----I-----NG-----41
DB	268	VDEHTINISLSLHHVEQMAIDMLTGNFYFVDIDDRIFVCNRRNGDTCVTLLDLELYNPKG 327
QY	42	AL-----VP-----P-----A-A-48
DB	328	IALDPAMGKVFYDYGQIPKVERCDMDGQNRKLVDSKIVFPHGITLDLVSLRVYWADAY 387
QY	49	--LI-----SIQKG-L-Q-Y--V-E-----A--EVS-I-----N 67
DB	388	LDYIEWVDYEGKGRQTIQ-GILIEHLYGLTFVFNLYATNSDNANAQOKTSVIRVNFN 446
QY	68	--E-----D--CTL-----72
DB	447	STEQVTVTRVDKGGALHIYHQRQRPRVRSHACENDQYKPGGCSDICLLANSHKARTCR 506
QY	73	---F---DG-----R-----PIE-----SL--81

Db	507	RSFSLGSDGCKKCKEHELFLVYKGRPGIIRGMDMGAKVDEHMIPIENLMNPAALDF	566	Db	1586	LYARQMEIRGVLDLAPYNYIISFTVPDIDNVTVLDYDAREQVYVSDVTRQAIKRAFNI	1645
QY	82	-----S-LI-----D-----AV-M-----PD--	90	QY	218	G-G-----QDVPSNK-----D-V-----TS-----LD-----	232
Db	567	HAETGIYFADTTSYLIHQKIDGTERETILKDGIIHNVEGVAVDMGMDNLYWTDGPKKT	626	Db	1646	GTGVETVSADLP-NAHGLAVDMVSRNLFWTSYDTNKKQINVARLDGSKFNAVOGLEBP	1704
QY	91	V-V-----QTR-----	97	QY	233	-----W-----N-S-----EG--TLL--A--TG-----S-Y-----	245
Db	627	ISVARLEKAAQTKTLIEGKMTHPRAIVVDPLNGMYWTDWBEEDPKDSRRGRLERAMMDG	686	Db	1705	HCLVHPHRLRGKLYWTDGNI SWANMDGSRNTLLPSGQKPGVGLAIDFPESKLYWISSGNH	1764
QY	98	-----A-Y-R-----D-K-----	102	QY	246	-----D-GF--A-R-----I-----W-----T-K-DG-----	256
Db	687	SHRDIPTVKTVLWPNGLSLDIPAGRLYWDAPYDRIETILLNGTDKIVYEGPELNHAF	746	Db	1765	TINRCLDGGLEVIDAMRSOLGKATALAIMGDKLWADQVSEKMGTCADGSGSVLVR	1824
QY	103	-----L-----AQQAA-----	109	QY	257	N--IA-----S--T--L--GO--	264
Db	747	GLCHGNLFWTEYRSGSVYRLRGVGGAPPTVLLRSRPPPIFEIRMYDAQQQVGTNK	806	Db	1825	NSTLWEMKMYDES IQLDHKGTNCPNVNNGDCSOLCLPTSTTRSCMCTAGYSRSGQ	1884
QY	110	-----AA-----A-----A-----A-RASQ--	120	QY	265	-----HKG-----P-----I-F-A-----L--	272
Db	807	CRVNGGSSLCLATPSRQCAEQVLDADGVTCLANPSYVPPQCPQGFACANSRC	866	Db	1885	ACEGVGSFLLYSVHEGIRGIPLDPNKSDALVPVSGTSLAVGIDFHAENDTIYVWDMGLS	1944
QY	121	Q-----G-----S-----A-----K-----N-G-----E	128	QY	273	-----K--W-----N-----K-----KGN-----F-I--L-----S	283
Db	867	IQRWKCDDGNDCLDNDSEAPALCHOHTCPSDRFKCENNRCPINRWLDCDNDGNSDE	926	Db	1945	TISRAKRDQWREDVVVTNGIGRVEGIAVDWIAAGNIYWTQDQFDVIEVARLNGSPRYVIS	2004
QY	129	N-T-A-----N-----G-----E-----E-----N	136	QY	284	AGVDK-----T-----T-----I--I--I-----	292
Db	927	SNATCSARTCPNPQFSCASGRCPISWTCDLDDCGDRSDESASCAYPFCPLTQFTCNN	986	Db	2005	QGLDPRAITVHPEKGYLFWTEWQYPIERRLDGTERRVLVNWISISWPNGISVDYQDG	2064
QY	137	G-----A-H-----T-----I-----	141	QY	293	--W-DA-----	296
Db	987	GRCININWRCDNDGDNDSDEAGCSHSCSSTQFKCNSGRCPPEHWTCTDGDNDGYSDE	1046	Db	2065	KLYWCARDTKIERIDLETGENREVVLSSNNMDFSVSFEDFIYWSDRTHANGSIKRG	2124
QY	142	AN--N-----HTD-----M-----M-E-----V-D--	152	QY	297	-----TG-----E-----AK-----OQ-----	303
Db	1047	THANCTNQATRPGGCHTDFQCELDGLCIPLRWRCGDGTDNDSSDEKSCGVTHVCDP	1106	Db	2125	KDNATDSVPLRTGIGVOLKDIKVFNRDRQKGTNVCANGGQCLYRGRGORACAH	2184
QY	153	G-D-----V-----E-I--PP-----NKAV-----V--	164	QY	304	-----F-----	304
Db	1107	SVRFCKDSARCISKAVWCDGNDGDNDSDENCESLACRPPSHPCANNTSVCLPPDKLC	1166	Db	2185	GMLAEDGASCREYAGVLLYSERTILKS IHLSDERNLNAPOFPEDPEHMKNVIALAFDYR	2244
QY	165	-----L-R-----G-H-----E-----S-----E-----V--	172	QY	305	-----P-----F--H-----S-----A-----	309
Db	1167	DGNDGSDGSEGEELCDQCSLNNGGCSHNCVAPGEGIVCPCPLGMLGPDNHTCQISY	1226	Db	2245	AGTSPGTPNRIFFSDIHFGNIQIINDGSRRTITVENUGSVEGLAYHRGWDLYWTSYTT	2304
QY	173	-----F-I-CA-----W-----N-----	178	QY	310	-----P-AL-----D-----V-D-----W-Q-----S	318
Db	1227	CAXHLKCSQCKDNKFSVKSCYEGWVLEPDGHSRSLDPFKPFIIFSNNRHEIRRIDLHK	1286	Db	2305	STITRHTVDQTRPGAFERETVITMSGDDHPRAFVLDECONLMFWTNNEQHPSIMRAALS	2364
QY	179	-----P-----V-----S-----D-----LL--A--S-----	186	QY	319	--N-----N-----T--F--AS-----C-----S-----	326
Db	1287	GDYSLVPEGLNRTIALDFHLSQALYWTDVVEDKIVRGKLLDNGALTSFEWVIQYGLATP	1346	Db	2365	GANVLTLIEKDIRTPNGLAIDHRAEKLYFSDATLKIERYDGHSHRYVLKSEPHVPG	2424
QY	187	-----G-----S-----GD-----S-----T--	192	QY	327	-----TD-----M-----C-I	331
Db	1347	EGLAVDWIAGNIYWNESNLQIEVAKLDGTLRTLLAGDIEHPRAIALDPRDGLFWTDW	1406	Db	2425	LAVYGEHFWTDWVRRAVQRANKHVGSNMKLLRVDPQOPMGIIAVANDTNSCELSPCRI	2484
QY	193	A--RI-----W-N-----	197	QY	332	-----HV--CKLG--OD-----R-----	340
Db	1407	DASIPIREANMSGAGRRTVHRETSGGWPNGLTVDYLEKRLWIDARSDAIYSARYDGS	1466	Db	2485	NNGGODLCLLTHQGHVNCSCR--GGRILQDDLTTCRAVNSSCRAQDEFCANGECINPSLT	2543
QY	198	-----LS-----E-----NST-----SG-----S-TQ--L	209	QY	341	-----P-----I--KT	344
Db	1467	GHMEVLRGHEFLSHPPFAVTLYGGEVYWDRTN--TLAKANKWTCHNVVTVVORTNQPFDL	1525	Db	2544	CDGVPHCKDSDEKPSYCNRRCKTFRQCSNGRCVSNMLWNCNGADDCGDSDEI PCNKT	2603
QY	210	V-----L-----R-----H-----C-----	214	QY	345	-----FQ-----	346
Db	1526	QVHPSPQMAPNPCEBANGGPGCSHLCLINYNRTVSCACPHLMKHLKONTTCYEPKFL	1585	Db	2604	ACGVGEFRCDGTICIGNSSRCNQFVCEADSDMNCSSATDCSSYFRLGVKGLFQPCERT	2663
QY	215	-----RE-----	217	QY	347	-----G-----H	348
				Db	2664	SLCYAPSWCDGANDCGYSDERDCPGVPRPCPLNYFACPSORCIPMSWTCDEKDCHEH	2723

```
QY 349 ---T-N---EV---N-A-I-K-W---D---PT---G 361
Db 2724 GEDTHCNKFCSEAQCQECQNHRCISQWLCDCGSDGDDCAAHCEGKTCGSPFSCFG 2783
QY 362 N-----L-----LAS-C--S--DD--M-----371
Db 2784 THVCVPERWLCDDKDCADGADESIAAGCLYNSTCDDREBFMCNQRCIQKHFVCDHDDC 2843
QY 372 -----T-----LK-----I--W-----S--NK 379
Db 2844 AGSDSEPCPEYTCGPSEFRANGCLSRQWECGENDCHDQSDAPKNPHCTSPHK 2903
QY 380 -----Q-----381
Db 2904 CNASSQFLCSSGRVAAELLCNQDDCGSDSDEBGRCHINECLSRKLSGCSQDCEDLKIGF 2963
QY 382 -----N-----CV-----H-----385
Db 2964 KCRCPGPRFKDGRTCADVDECSCTTFPCSQRCINTHGSYKCLCVBGAAPRGDPSCKA 3023
QY 386 -----D-----L-Q-----Q-----389
Db 3024 VTDEBFLPANRYLRKLNLDGNTLLKQGLNNAVALDFDYREQMIYWTVDVTTQGSMI 3083
QY 390 -----H-----N-----K-----EI-----Y-TI-----397
Db 3084 RMHLGNSNVQVLRHTGLSPDGLAVDWVGGNLYWCDCGRDTIEVSKLNGARTVLVSSG 3143
QY 398 --K-----W-----S-----P-----T-----402
Db 3144 LREPRALVVDVQNGYLTWTDGSHSLIGRIGMGSSRSVVDTKITWPNGLTLDVYTERI 3203
QY 403 -----G-----P-----CT 406
Db 3204 YWADAREDYEFASLDGSRNHRVLSQDIPHIFALTLEFEDVYVYWTWETKSIINRAHKTGT 3263
QY 407 N-----N-P-----N--N--ANL-MLAS-----A-S-P- 420
Db 3264 NKTLLISTLHRPMDLHVHALRQDPVPHPCVKNNGCCNLCLL-SPGGGHKCACTPNFY 3322
QY 421 ---D-----S--T---V---R---LW---DV-----D-----R-----431
Db 3323 LGSGRTCVSNCTASQFVCKNDKCIPIFWKDKTDEDCGDSBDEPPDCPEFKCRPGQFQCS 3382
QY 432 -GIC-----IHT-L-----T-K-H-----Q-----442
Db 3383 TGICTNPATFICDGNDCQNSDEANCIDHVCPLPSQFKCTNTRCIPGIFRCNGQDNCGDG 3442
QY 443 -----E-----P-VY-----S-----V-AF---450
Db 3443 EDERDCPEVTCAPNQFQCSITKRCIPRVWVCDRDNDVDSDEPANCNTQMTGVDPEFCK 3502
QY 451 -S-----P-----DG-----R---Y-----LAS--G 460
Db 3503 DSGRCIPARWKCDGEDDCGSDGDEPKECDERTCEPYQFRCKNRCVPGRWCDVNDGCG 3562
QY 461 -----S-F-----D-K-CV-----466
Db 3563 DNSDEESCTPRPCSESEFSCANGRCIAGRWKCDGHDHCDGSDKCDKTPRCMDQFQCKS 3622
QY 467 -H-I--W-----NT-----471
Db 3623 GHCIPLRWKCDADACMDGSDDEACGTGVRTPLDFEQCNLTCKPLAWKCDGEDDCGN 3682
QY 472 -----QVCL-----H---476
Db 3683 SDENPEECARFVCPNRPFRCKNDRVCLWIGRCQDGTDCGDTDECEPPTAHTTHCK 3742
QY 477 ---YL-----N-----479
Db 3743 DKKEFLCRNQRCLSSLRNMFDDCGDGSDEEDCSIDPKLTSCATNASTICGDEARCVTE 3802
```

```
QY 480 -----GQ-----V---L-----L--NL--G-----R---S---IC-- 491
Db 3803 KAAACACRSRGFTVPQPGCQDINECLRFGTCSQLCNTKGGHLCSARNFMKTHNTCKA 3862
QY 492 -----LY-----493
Db 3863 EGSEYOVLYIADNDNEIRSLFPGPHSAYBQAFQGDSEVRIDAMDVHVKAGRVVYTNWHTG 3922
QY 494 T-----LP-----H-----HL-----V-----VI 502
Db 3923 TTSYRSRSLPAAAPTTSNRHRRQIDRCVTHLNISGLKMPRGIAIDWAGNVYWTDSGRDVI 3982
QY 503 -----P-LVA--L-----IE-----L--LV-- 512
Db 3983 EVAQMKGENRKTLSGMDIDEPHALVVDPLRGTMWSDGWNHPKPIETAAMDGTRELIVQD 4042
QY 513 -----L-----K 514
Db 4043 NIQWPTGLAVDYHNERLYWADAK 4065
```

RESULT 13

A53102

alpha-2-macroglobulin receptor precursor - chicken

N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor;
C:Species: Gallus gallus (Chicken)

C>Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 09-Jul-2004

C:Accession: A53102

R:Nimph, J.; Stifani, S.; Bilous, P.T.; Schneider, W.J.

J. Biol. Chem. 269, 212-219, 1994

A:Title: The somatic cell-specific low density lipoprotein receptor-related protein of t

A:Reference number: A53102; MUID:94103212; PMID:7506255

A:Accession: A53102

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-4543 <NIN>

A:Cross-references: UNIPROT:P98157; GB:X74904; NID:G438006; PIDN:CAAS2870.1; PID:G438007

C:Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated

C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding

C:Keywords: beta-hydroxyaspartic acid; beta-hydroxyaspartic acid; calcium binding; glycopro

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-3942/Domain: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>

F:29-66/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:74-110/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:117-150/Domain: EGF homology <EG1>

F:156-190/Domain: EGF homology <EG2>

F:200-241/Domain: LDL receptor YWTD-containing repeat homology <YW01>

F:242-283/Domain: LDL receptor YWTD-containing repeat homology <YW02>

F:294-336/Domain: LDL receptor YWTD-containing repeat homology <YW03>

F:337-380/Domain: LDL receptor YWTD-containing repeat homology <YW04>

F:381-422/Domain: LDL receptor YWTD-containing repeat homology <YW05>

F:423-470/Domain: LDL receptor YWTD-containing repeat homology <YW06>

F:480-521/Domain: EGF homology <EG3>

F:573-615/Domain: LDL receptor YWTD-containing repeat homology <YW07>

F:616-661/Domain: LDL receptor YWTD-containing repeat homology <YW08>

F:662-712/Domain: LDL receptor YWTD-containing repeat homology <YW09>

F:713-754/Domain: LDL receptor YWTD-containing repeat homology <YW10>

F:755-797/Domain: LDL receptor YWTD-containing repeat homology <YW11>

F:805-840/Domain: EGF homology <EG4>

F:852-888/Domain: LDL receptor ligand-binding repeat homology <LDL3>

F:893-929/Domain: LDL receptor ligand-binding repeat homology <LDL4>

F:934-969/Domain: LDL receptor ligand-binding repeat homology <LDL5>

F:974-1009/Domain: LDL receptor ligand-binding repeat homology <LDL6>

F:1013-1049/Domain: LDL receptor ligand-binding repeat homology <LDL7>

F:1060-1095/Domain: LDL receptor ligand-binding repeat homology <LDL8>

F:1102-1138/Domain: LDL receptor ligand-binding repeat homology <LDL9>

F:1143-1180/Domain: LDL receptor ligand-binding repeat homology <LDL10>

F:1183-1219/Domain: EGF homology <EG5>

F:1225-1259/Domain: EGF homology <EG6>

F:1267-1306/Domain: LDL receptor YWTD-containing repeat homology <YW12>

F:1307-1353/Domain: LDL receptor YWTD-containing repeat homology <YW13>

Db 3116 CDKGRDITVSKLNGAIVRTLVNSGLRPREALVVDVQVNGVLYWTDWGDHSLGKIGMDGT 3175
QY 400 --S-----P---T-----G-----P----- 404
Db 3176 NRSVIVDTKITWPNGLTLDYINSRIYWADAREDIYEFASLDGSGNRHTVLSQDIPHFALT 3235
QY 405 -----GTN-----N-P-----N 410
Db 3236 LFEDYIYTDWETKSINRAHKTITGANKTLLISTLRPMDIHIVHPYQPDVPHNPKCTNN 3295
QY 411 A---NL-MLAS-----A-S-F---D-----S-T---V---R---LW---DV--- 429
Db 3296 AGCSNLCLL-SPGGGHKACPTNLYLSDGKTCVSNCTASQFVCKNDKICPFWKKDTE 3354
QY 430 ---DR---GIC-----IHT-L--- 438
Db 3355 DCDGRSDEPDCPEFKRPFQFQCSTGICTNPAFICDGNDCQDNDSEANCDIHVCLPSQ 3414
QY 439 ---T-K-H---Q-----E-----P-VY----- 446
Db 3415 FKCTNTRCIPGIFRCNGQDNCGGDEKDCPEVTCAPNQFQCAITKRCIPRVWVCDRDN 3474
QY 447 ---S-----V-AF---S---P-----DG-----R--- 455
Db 3475 DCDVDSDEPANCQTMTGVDPEFRCKDSGRICIPARWKCDGEDDCDGDGDEPKECBERTCE 3534
QY 456 -Y-----LAS-G-----S-F----- 462
Db 3535 PYQFRCKNNRCVPRWQCDYDNCDSNDEBCTPRPCESEFSCANGRCIAGRWKCDGD 3594
QY 463 ---D-K-CV-----H-I---W----- 469
Db 3595 HDCADGSDKDCIPRCEFDQYQCKNGHCIPMRWRCDADCDMDGTDEDCGTGVRTCP 3654
QY 470 ---NT-----Q-----VCL----- 475
Db 3655 EFCQNTLRKPLAWKCDGEDDCDGNSENDEPECLKQCPNRPFRCKNDRVCLWIGRQCD 3714
QY 476 -----N-----G-----QVLL-----N-L-----G 487
Db 3835 LCNNTKGVSHVCSCAKNFMKTDNMCKAEGSRHQILYIADDNKIRSMYFPNPNAYEPAFQ 3894
QY 488 ---R-----S-I-C----- 491
Db 3895 DENVRIDAMDIIYVGNKIYTNWHTGRISYCELPASSAASTASNNRRRQIDGGVTHLNIS 3954
QY 492 -L-----Y-----TL-----PH----- 497
Db 3955 GLKMPRGIAVDWAGNIYWTDSGRDVEIVAQMGKGNKRLISGMIDEPHAIIVDPLRGTM 4014
QY 498 ---H-----LV-----VI----- 502
Db 4015 YMSDGNHNPKIETAAMDGTILRETLVODNIQWPTGLAVDYHNERLYWADAKLSVIGSIRLN 4074
QY 503 ---PLVA-----L-----I-E-----L-----L----- 511
Db 4075 GTPVPAIDNKKGLSHPFSDIFEDYIYGVYTYNNRIFKHKHGKSVTNLTSGLNHATD 4134
QY 512 -VL-----K 514
Db 4135 VLYHQYK 4142

RESULT 14
A47437

LDL-receptor-related protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A47437; S27801; T21547
R:Yochem, J.; Greenwald, I.
A:Title: A gene for a low density lipoprotein receptor-related protein in the nematode C. elegans. Natl. Acad. Sci. U.S.A. 90, 4572-4576, 1993
A:Reference number: A47437; MUID:93281621; PMID:8506301
A:Accession: A47437
A:Molecule type: DNA
A:Residues: 1-4753 <YOC>
A:Cross-references: UNIPROT:Q04833; GB:M96150; NID:G156359; PIDN:AAA28105.1; PID:G156360
A>Note: nucleotide sequence not given; translation not complete in this paper
R:Yochem, J.; Greenwald, I.
Submitted to the EMBL Data Library, July 1992
A:Description: A gene for an LDL receptor-related protein (LPR) in the nematode C. elegans
A:Reference number: S27801
A:Accession: S27801
A:Molecule type: DNA
A:Residues: 1-4753 <YOC>
A:Cross-references: EMBL:M96150; NID:G156359; PIDN:AAA28105.1; PID:G156360
R:Wilkinson, J.
Submitted to the EMBL Data Library, June 1996
A:Reference number: Z19439
A:Accession: T21547
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4753 <WIL>
A:Cross-references: EMBL:273907; PIDN:CAA98124.1; GSPDB:GN00019; CESP:F29D11.1
A:Experimental source: clone F29D11
C:Genetics:
A:Gene: LPR
A:Map position: 1
A:Introns: 3/1; 88/1; 132/1; 172/3; 219/1; 298/1; 463/2; 526/2; 585/3; 780/2; 874/2; 975/15/1
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
C:Keywords: tandem repeat; transmembrane protein
F:53-87/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:92-131/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:138-175/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:182-218/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:223-257/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:262-297/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:302-336/Domain: EGF homology <EGF1>
F:1054-1095/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:1101-1138/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F:1146-1182/Domain: LDL receptor ligand-binding repeat homology <LDL9>
F:1187-1223/Domain: LDL receptor ligand-binding repeat homology <LDL10>
F:1228-1263/Domain: LDL receptor ligand-binding repeat homology <LDL11>
F:1270-1307/Domain: LDL receptor ligand-binding repeat homology <LDL12>
F:1313-1350/Domain: LDL receptor ligand-binding repeat homology <LDL13>
F:1359-1396/Domain: LDL receptor ligand-binding repeat homology <LDL14>
F:1441-1475/Domain: EGF homology <EGF>
F:1611-1654/Domain: LDL receptor YWTD-containing repeat homology <YW33>
F:2792-2829/Domain: LDL receptor ligand-binding repeat homology <LDL15>
F:2834-2868/Domain: LDL receptor ligand-binding repeat homology <LDL16>
F:2874-2912/Domain: LDL receptor ligand-binding repeat homology <LDL17>
F:2919-2956/Domain: LDL receptor ligand-binding repeat homology <LDL18>
F:2961-2997/Domain: LDL receptor ligand-binding repeat homology <LDL19>
F:3006-3044/Domain: LDL receptor ligand-binding repeat homology <LDL20>
F:3049-3093/Domain: LDL receptor ligand-binding repeat homology <LDL21>
F:3100-3135/Domain: LDL receptor ligand-binding repeat homology <LDL22>
F:3140-3174/Domain: LDL receptor ligand-binding repeat homology <LDL23>
F:3187-3222/Domain: LDL receptor ligand-binding repeat homology <LDL24>
F:3586-3623/Domain: EGF homology <EGF1>
F:3627-3666/Domain: LDL receptor ligand-binding repeat homology <LDL25>
F:3671-3705/Domain: LDL receptor ligand-binding repeat homology <LDL26>
F:3709-3746/Domain: LDL receptor ligand-binding repeat homology <LDL27>
F:3753-3788/Domain: LDL receptor ligand-binding repeat homology <LDL28>
F:3793-3830/Domain: LDL receptor ligand-binding repeat homology <LDL29>
F:3833-3871/Domain: LDL receptor ligand-binding repeat homology <LDL30>
F:3878-3912/Domain: LDL receptor ligand-binding repeat homology <LDL31>
F:3917-3951/Domain: LDL receptor ligand-binding repeat homology <LDL32>

F:3959-3995/Domain: LDL receptor ligand-binding repeat homology <LD33>
F:4000-4040/Domain: LDL receptor ligand-binding repeat homology <LD34>
F:4049-4083/Domain: LDL receptor ligand-binding repeat homology <LD35>
F:4092-4130/Domain: EGF homology <EGF2>
F:4343-4386/Domain: LDL receptor WYTD-containing repeat homology <YW38>

Query Match 70.0%; Score 2553.7; DB 1; Length 4753;
Best Local Similarity 9.8%; Pred. No. 3.7e-52;
Matches 428; Conservative 70; Mismatches 10; Indels 3853; Gaps 375;

QY 1 MSIS-----SDE-----V-----N-PL-----V-----12
Db 105 HVSCTAEYORCDNITDCADGSDKCPASTVDCSSQNVFMCADGRCQCFVSKKCDGKYD 164
QY 13-----Y-----13
Db 165 CRDLSEKDCSRNHTACQYQPCADKTQCIQKSWVCDGSKDCADGSDPDTCEPKCT 224
QY 14-----R-----Y-----15
Db 225 ANEFQCKNRCQPRKFRCDYDDCGSDSEDECGEYRCPPGKWCNCPGTGHCIDQLKCDG 284
QY 16-----L-Q-----E-----SG-----FSH-----S-----24
Db 285 SKDCADGAEQCSQNLCPSLGCAQCHPSHGECTCFSGYKLDLDRF-HRTCSDINECA 343
QY 25-----A-----FT-----27
Db 344 EFGYCDQLCANHRPGFTCSCLGDCFTLOMEHGGKDNLTWRGVCVSNADKMLFVARRE 403
QY 28-----P-GI-----K-----SH-----ISQ-----36
Db 404 GLYRLNPKNPDEVKKLASGEFTYGDYDGRKIEFTWDLRAHSAFSDVDEGEISQIK 463
QY 37-----S-----N-----I-----39
Db 464 KLSLSLVPRCLAVDWTNTHYIESSRRIDVSSYDGERRTVLLADGLTLPDLALDP 523
QY 40-----N-----G-----A-----L-----43
Db 524 LRGEFTNQLKLEAAMDGTNRRLTVTHQVSGIVVDITAKRVYVWDPKVDRLSD 583
QY 44-----VP-P-A-----A-L-I-----S-----I 52
Db 584 YQGNDRRIVAQGNVVPHPFGLALPDQLYWTDWTRLGVIQVEKFGSDTKLLWSNTENN 643
QY 53-----I-----Q-----K-----GL-----Q 58
Db 644 FPMGISAYHPMAQPGQSECLAMKIENFCTNADCEGMCILSKDNGGFGVYKCAPIGQ 703
QY 59-----Y-----V-----EA-----EV-----64
Db 704 KLVNGKRCIDSIDYLLFPSSNKIVRGIFPEINEKALAEAVLPISQRRIGMYFEVCDV 763
QY 65-----S-I-N-----E-----DG-----70
Db 764 HGNSFFYADIMNTIYIRPDGGAAPVLVTHDGLFSMSFDWISKQLYYVDNIRNSLEV 823
QY 71-----T-----LP-----D-----74
Db 824 VKIGETGLVHPDELVRQLITELRDPVSVVHPWKGLLFYAEAMRPAAYRCHIDQNCQ 883
QY 75-----GRP-I-E-----SL-S-----L-ID-A-VM--88
Db 884 VIRNTLGRPSEMAIDFAENRLCWGOTLLKTISCMDPFGKGVVKKLDINPIVAITIME 943
QY 89-----P-----D-V-----V-----92
Db 944 YIYVYHQPYSIRVHKNGGSGKIYREFGADERSIFSLKACSHQNPDPDSREHPCRA 1003
QY 93-----Q-T-----RQ-----Q-A-----Y--99
Db 1004 SOCTQLCFATPSHSHNELEAKCACRQGFWMINKENNHSQKDPAEKIEQLCSNSNQFQC 1063

QY 100 ---R-----DK-----L-AQ-----105
Db 1064 KNGRCIPKWKCDGENDCLDESDEIDKDGKCFHETECAENTIKCRNTKKCIPAQYCGDG 1123
QY 106 -----Q-----Q-----107
Db 1124 DNDGSDYDESDYKCKGQKPVCAAKKFOCDNHRCIPEQWKCDSDNDGSDGSEKLEMC 1183
QY 108 -A-AA-----A-----A-----A-----113
Db 1184 NATCAANQSCANGRCIPIYWLCDGNDYDGTDEDKERCPPVQCSALQFCANGRCQVCP 1243
QY 114 -----AAAA-----AS-----QO-----GS-----123
Db 1244 LRNHCDGSDGSDGSDSCAVTAESCTPDQFKCVSSGLCIPASWKCDGQDQDCGSD 1303
QY 124 -----A-K-----N-----GEN-----T-----130
Db 1304 KFGCTSGROCSDQFKCGNGRCILANNWLCGENDGSDGSDSERGCKTSMNARKCPFEH 1363
QY 131 -A-N-----GE-----E-----NG-----137
Db 1364 VACENDQETCIPHLQLCDGKTHCPGTDGGRCARDLCSADRAGCSFKCHNSNGPICSC 1423
QY 138 -----AH-----T-IA-----142
Db 1424 PFGEQLVNTKCEPENECILSSSSQRCQKDEKHGFTCSDEGYELVDVKRTCKVADNVK 1483
QY 143 -----N-----N-HT-----DMWE-----150
Db 1484 TRIYVSNRRIYSDHKLDNWHTFGAIVENAJALAWDSLTDRIYSDIREKKILSANRNG 1543
QY 151 ---V---DG-DV-E-----I-----P--158
Db 1544 TNATFIADGLDITGEGIALDWGRNLYWVDSNLTIEVANLEDPKORTLLVHQVNSQPRG 1603
QY 159 ---P---N---KA-----VV-----L---R-----166
Db 1604 IAVDPKGMVFWTDNGQNCPIERASMDGTDRQIIVKTIYWPNTIALDYTTDRVYFADSK 1663
QY 167 -----G-----H-----E-----SE 171
Db 1664 LDFIDFVNYDGSRTQVLASSKFVQHPHALAIFEDMMYISDRRLQKLVYKPYKNGTTSE 1723
QY 172 ---VF-----I-----174
Db 1724 YPSHTFSKALGVVAVHPVLQPVIKNNPCSTNPCSHLCLLNKNKNTFTCKCPMGEKLDASGK 1783
QY 175 -C---A-----W---N-----PV-----S-----D-----182
Db 1784 KCIDDAKPLVIIQKTNVFGIEMNSASEKETPVLAGMVPLSLGLGNADFADAYDALSEEMFI 1843
QY 183 L-----LA-----S-----G-----G-D-----190
Db 1844 LEHTNHAKTLAGIQTDSAIYRSTVNGNKTMPSSAVPDADAYCLGFDWNGRNLVVGNKIT 1903
QY 191 -S-----TA---R-----I-----W-----196
Db 1904 QTIEIRTOGKQYRSVILNSDOSPTAVTVPVIAVDADKGYVFWLDRGGGAADAKVARAG 1963
QY 197 ---N-L-S-E-----N-----S-T-----S-----204
Db 1964 LDGSPNPLVIASNDLAEHLDAIDTTNTRYVFSKAGRAGISSVYDQDRHYVLSDCGRQP 2023
QY 205 -----G-----S-----T-----Q-----L--V--L--R-----212
Db 2024 NGLAFYGRDLRFYADSAFDSIEVATINGDSQPPQWTHFKKDVENLANIKALQPRASSGHP 2083
QY 213 -----H-CI-----R---E-G---CO-----D--221
Db 2084 CHINNGNCNDHICIPLMFAORTCTCANGYVKDQGTCKLFDSEFVIVATKTKVIGYPIDET 2143

QY 222 ---V---P---S---NK---D-V--- 228
Db 2144 QSKGVAMEPIGGLSITGVYDYBSKTYIYVAESGINKGITAYIGESSPRAVIRDSIGSL 2203
QY 229 T--SL--DW---N---S--E---GT---LL--- 240
Db 2204 TIKSLAIDWINYMYFINHAERTNIEVSKLDGYRKILLTTKTETPSSIAVDVPSRYLY 2263
QY 241 -A---T---GS---Y---DG--- 247
Db 2264 WADQOGKPTTQSRFLDGRREVIVSSGIAEPTDLVVDVASKMYIWSDAKMDGIYRVRSTG 2323
QY 248 ---FAR--I---WT---KD-G 256
Db 2324 GTPELVRSDIASAAGVALHGQNMWTDNRLKLEKFRATSKPNQTSLLSPPTVAASLKDIG 2383
QY 257 ---NL--- 258
Db 2384 DVAVFSSNNQPRASSPCQITDNLKSPCTQLCFATPCTQTPTCSCARGVLKGRTCBEPDT 2443
QY 259 ---AS---T--- 261
Db 2444 YIMFSDGDKIIDVAIBPDVKASRPLKDPPEISNLQTFDVDVNLRRVYFVESPVGWIS 2503
QY 262 ---L-G---Q---H- 265
Db 2504 WFSMNAENPRLVFGASKOPHAKKEIRHISDMKLDWLTQKIYFTTGRGKVMADTAGHEHL 2563
QY 266 ---F---A---L---K---W---I--- 269
Db 2564 STIASGDWTYALAI DPCSGLLFWSDGYKTSGLGYEPRIERSNLAGSRKVI VSESISLP 2623
QY 270 ---F---A---L---K---W--- 274
Db 2624 AAI AVDFRNQKIY WADVNLNIEVADYDGNRKRVIASGYRAKSLDIWDRWLYMSDPLSNG 2683
QY 275 ---N--- 275
Db 2684 VFRIDKESGLENVSDRRIPGALRVFASESDVTRNVQCNALTSOLCKTDNGGCDQLC 2743
QY 276 ---K---K---G---NF--- 280
Db 2744 TVVADDIGLAASKVQCSNDTYELVQEPGKYPTQCVLRGNSSEPAKECLPPYFQCGDG 2803
QY 281 --IL--- 282
Db 2804 SCILLGATCDSPKPCADASENPYCNTRSCPEDYNLCTNRRCIDSAKKNHIDDCGDS 2863
QY 283 ---SA---G---V-DK---T--- 289
Db 2864 DELDCPSAVACABGTFFCSNGHCINQTKVCDGHNDCHDEQVDESIALTCPLPIDCRGVK 2923
QY 290 ---T---I--- 292
Db 2924 VRCPTNINI CIQPADLCGYDDCKADENQLFCMNOQCAQHYVRCPSGRICIPETWQCDGD 2983
QY 293 ---WD-AHT---GEAK---Q---F--- 304
Db 2984 NDCSGDWDETHNTCTDAGK-KICVGDYLFQCDNLKICISRAFTICDGEDDCGDSDEHSRH 3042
QY 305 ---PRH--S---A---P-A-L---DV---D--- 315
Db 3043 GCGNRTCTQEFHCTNAKLAQPKYECIPRAWLDCDGDVTCAGDEBSELCKTEKKECNK 3102
QY 316 ---W---Q--- 317
Db 3103 GEFRCSNQHCISTWECDDGNDCLDGDSEHANCYSSCQDPDPQCANHKVNPWSKCDGN 3162
QY 318 ---S---N---N---T---F---AS--- 324
Db 3163 DDCEDGSEDKCPKNSASQAQASKCSNGQPQCTSGECIDDAKYCDRNFDCDRSDESLC 3222
QY 325 ---CS---TDM---C---I---H-V---C- 334

Db 3223 FIDECSLAEPLCEQKCMKIGYKCDCEGFAIDISDQKSCHNVNCEYEGISGCSQKCD 3282
QY 335 -KLG---Q---D---R---P---I-K-T---F- 345
Db 3283 DKIGSYKCGVDGYQLSSDDHSCRKTEMEPEFFLLANKHYIRKISIDGNKYELAAQGF 3342
QY 346 ---OG---HTN---E---V- 352
Db 3343 NVVSLDIDLTEKAYLIDQGLRLRLRVDLDEMDSPLSSYETVLRH-NVYTEGIAVDWVG 3401
QY 353 ---N---AI---K---W--- 357
Db 3402 RKLMLNROERSIRVCELDGRFCKTLIRDRIOQPKAIVVHPGKGYLFFTEWSLOPVIGRI 3461
QY 358 ---DP---TG-N---L---A 365
Db 3462 ALDGPSELQDPIFKLAHDLGWPNAIAIDYFSDRLFWGDAHLNEIGFMDPDGNGRRHIPA 3521
QY 366 ---SCS---DD-M---TL-KI---WS---MKQ- 380
Db 3522 QRTSHVS-SMVVFDYLYWADNWLREVLRCKWTGKNETILKTVLPNDLRIVHPMRQP 3580
QY 381 ---DN---CVH---DLQ---Q---HN-KE-IYT--- 396
Db 3581 AYPNPGDNNGCSHCLIGAGNGYTCSPD-QFVLLSDQKTCENCTERQFACGGDDA 3639
QY 397 --I-K-W---S-P-T-G---P-G---TN-N---P---N- 410
Db 3640 KCIPLKMYCDGBCRDSDEPESICGQRI CPVGBFQCTNHCTRPFOICDGNDDCGDS 3699
QY 411 ---ANL---ML---A---S-A---S---F- 420
Db 3700 SDEQNCDAKCPMWFKAATGRICIPRFTCDGDDCGDRSDEADTLCMSAERNCTAEFR 3759
QY 421 ---D-S--T---V---R---L---W--- 427
Db 3760 CANNKCIKAWCRDNDGCGSDETPCAQIECKGWTRCSCSSVRCIPNNAFCNGQDDC 3819
QY 428 ---DV---D--- 430
Db 3820 RDNSEDKQRCPTCDDVGEFRCATSGKIPRRWMCMTENDCGDNSDELASCGTTRPCS 3879
QY 431 ---R-G-I---C-I---H---T-LTKH---QE 443
Db 3880 ESEFRNDGKCI PGSKVCDGTIQCSDGLDESQTLRRCLPGRHQCDGTGICIAHKWCDRK 3939
QY 444 ---P---VVS---VAF---S-P---DG--- 454
Db 3940 KDCPNADELHCEDV-SRRTCSPPFEPCANSVICIPKFMCDGNDGDSNDSSESRSA 3998
QY 455 ---RY-LA-S---G---SF---D--- 463
Db 3999 QCDPPLRFRCAHSRLCNILQLCNGFNDGPNDFSDHLSMCSSEFSEYDGCSSDQFKCAN 4058
QY 464 -KCV---HI---W--- 469
Db 4059 GKCVNGTVACDRKDCGDASDEIGCSKHGKTSCEAFNGNGGCKHICTDVRDGFYCHCRD 4118
QY 470 ---N-TQ---V---C--- 474
Db 4119 GFRPDQSPKECIDDECAGNNTCTQLCLNTKGYLCRCHEDYNNVVGSMTGKDCRAK 4178
QY 475 ---LH---Y--- 477
Db 4179 GDAANVMIGADSLVQLSLHSGTNRHAAKANDDDNDIIGIAFDPKELMTYWDGERT 4238
QY 478 -L---N---GO---V---LL---NL-G--- 487
Db 4239 IYRSIAANGSHGQKLDVDFAAMGVFTAIADVTTGNLFLIAAVSENIENGLVTRAKK 4298
QY 488 R-S---I-CL---Y---T-L--P---H--- 497

Db 4299 RMSEPIDNQTGFVCLPDGRYLKXIVAGHLQOPTALTAPAGRICYSDAGLHAKIEC 4358
QY 498 -----H-----L-----E-----V-----V-----501
Db 4359 ADMGTHRQIIIVKDLVFSPTSMADSGKGNRIYVDPKRYRVDVAVNDGSRITVVVDRH 4418
QY 502 IPL-V-AL-----I-----E-----LLV-----LK 514
Db 4419 IPYAVDFVFNHXYWLSRESKTLVQDKFGRGRVSVLASDLE 4459

RESULT 15
T42215
N:Alternates - mouse
N:Alternate names: sperm-specific membrane protein
C:Species: Mus musculus (house mouse)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T42215
R:Gao, Z.; Garbers, D.L.
J. Biol. Chem. 273, 3415-3421, 1998
A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane protein
A:Reference number: Z22080; MUID:98123114; PMID:9452463
A:Accession: T42215
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5376 <GAO>
A:Cross-references: UNIPROT:O88799; EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC266
C:Genetics:
A:Gene: Zan
A:Map position: 5
C:Function:
A:Description: functions in multiple cell adhesion processes
A:Note: found exclusively on the apical region of the sperm head
C:Keywords: cell adhesion

Query Match 70.0%; Score 2552.7; DB 2; Length 5376;
Best Local Similarity 10.1%; Pred. No. 5e-52;
Matches 427; Conservative 68; Mismatches 14; Indels 3713; Gaps 388;

QY 1 M-S-----S-----D-----E-----V-N-FL-----V 12
Db 1156 MFSLAPTTGVSTTESCPNNAHIELCACPASCSFKPSKPCPPICPGVCNPGFLFSNNQCI 1215
QY 13 -----Y-----R-----S-----H-----YLO- 17
Db 1216 NESSCNCYNNKHYPGEBEFTNCTERCCLPGSLMECQISQCGTHTVCOLKSDQY-OC 1274
QY 18 E-----S-----GF-----S-----H-----23
Db 1275 EPYKATCLVYGLHFVTFDERHIGFTGTCTYILTQTCNSTDHPFRITANTEERGVEGV 1334
QY 24 -----S-----S-----A-----F-----TFG-----29
Db 1335 SCLDKVVISLPETVTMISGRHTLIGDQVETLPAILSDDDTYVGLSGRFVELRTTFGLVR 1394
QY 30 -----I-----KS-----H-----IS-Q-----36
Db 1395 WGDQDLFTVTSFSGKLGCGNVDGSSNDNLKSDGMTHDEELRLSNQVDEDEK 1454
QY 37 -----S-----N-----IN-G-----A-----42
Db 1455 DWVSSRCQKKPPSCDAALGSTMSGKLGCLGVNPSGPFPEACLLHLKASSFLDNCVTDM 1514
QY 43 -----L-----V-P-----PA-----A-----48
Db 1515 CSFQGLQQLCARMSAMTATODAGYVPKWPREPQFCPLVCPKNSRYSILCAKPCPETCHP 1574
QY 49 -----LI-----S-----S-----I-----52
Db 1575 ISTQHCSKCEGCEBDFGLSGSECVSPSQCGCTSFQGRYFKLQEQWFNPDCKEIT 1634
QY 53 -----I-----Q-----K-G-L-----Q-----58

Db 1635 CESHNHLCKPMKCAQEACSYKNGYLGCHGAQAATCMVSGDPHYLTDFGALHFMGTCT 1694
QY 59 YV-----E-AB-----VS-INE-DGTL-----PD-----G-----R 76
Db 1695 YVLTQPCWKSQBNFVWSATNEIHDGNLEVSVKAVHVQVFDLKLKSMFKQKVLLNQR 1754
QY 77 -----P-----I-----E-----SL-----81
Db 1755 VVLPVWPSQGRVTRILSGIFVLLYTNFGLQVRYDGRHLVEVTPVSSYTGSLCGLGNYNN 1814
QY 82 -----SL-----I-----D-----AV- 87
Db 1815 NSMDDNLRADMRPAGNSLLGAAWKILEASDFCFLAGGKPSRCADSDMDVWTKCAIL 1874
QY 88 M-----P-----DV-----V-Q-----T-----R-----Q-Q-----97
Db 1875 MNPGLPFSNCHRAVPPQASFCVYQCCTNGDNLTFCHSLQAYASLCAQAGQVTTWRNS 1934
QY 98 -----AY-----RD-K-L-A-----Q-----Q 106
Db 1935 TFCPMRCPPRSSYNPCANSCPATCLTLSTPRDCPTLPCVEGCECQSHILSGTTCTVPLRQ 1994
QY 107 -----Q-----AA-----AA-A-----AA- 114
Db 1995 CGCSDQDGSYHLLGBSWYTEKTCTTLCTCSAHSNITCPTACKANHVCLRQEGLLRCAE 2054
QY 115 -----AAAA-----S- 119
Db 2055 MGECRISDSQIVSFDHSHPIQDCTYILVKVCHPNTNMPFMISAKTDINTNGKNKTF 2114
QY 120 -----Q-----G-----122
Db 2115 GYVQLYIDIFNHITLQKDLHLVLSINDSIVTLPTTHIPGVSVMTEDVYTVITIKDEI 2174
QY 123 -----S-----A-----K-----NGE-----N-----TA 131
Db 2175 QVKFESNNFLDKVIPASSNGKVCVCGNFGNGBEDELMTPSGELARDEOEFMNSWKDKSM 2234
QY 132 --N--G--E--E--NG--AH--T--I--141
Db 2235 DPNCKIEQNLQVQOEIMNGKCRIPIDPEKAQANCQTALQGPAMAHCSRVPDKPFLK 2294
QY 142 -----A-----N-----N-----H-T-----146
Db 2295 CMNSCFEPRELFRALCDSLSQSFEDACONQGLKPPWRNSFCLECPAHSHYTNCLPSCP 2354
QY 147 -----D-----MM-----E-----V-----D-----G-D-V-----155
Db 2355 PSCLDPDSSRCGSGHKVPATCREGTCQPDYVLLNDKCVLRSHCGCKDAQGVIPAGKTW 2414
QY 156 --E--PP--159
Db 2415 ISEDCQSTCMKGMRCWDFQCPPTGYCKNSNDSSNCVKISLQCPAHSKFTDCLPPCH 2474
QY 160 -----N-K-----A-----V-----VLR--G--H-----168
Db 2475 PSCSDPDHCEGISNNAHSCGCVQGYVLRNDKCVLRIECGQHTQGGFIPAGKWW 2534
QY 169 -----E-----S-----E-----V-----F-----173
Db 2535 TSGCSQSCDCMEGVIRCNQFQCPSTGYCQDIEDGTSNCANITLQCPAHSSFTNCLPPCQ 2594
QY 174 -----I-----C-----A--W 177
Db 2595 PSCSDPEGHCGSGTTKAPSACQEGCVCEPDYVVLNNKCVPRIECGCKDAQGVILPADKWW 2654
QY 178 -N--P--V--S--D-----L 183
Db 2655 INKGCTQTCACVTGTHCEDFQCPSTGYCKDIDKDDASNTEILLQCPDHSLYTHCLPSC 2714
QY 184 LA-S-----G-----S-----G-----D-----ST--A-RI-----W 196
Db 2715 LSCSDPDGLCRGTSPEAPSTCKEGCVCDPDYVLSNDKCVLRIECGCKDAQGVILPAGKTW 2774

QY 197 -----NL--SE-----NS--TS----- 204
Db 2775 INRGTSQSCMGGAIOQNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 2834
QY 205 -----GS-----TQ--LV----- 210
Db 2835 PSCSDPDGHCESSTKAPACKGVCCEPDYVYLNKNCVPRIECCGCKDTQGVLIIPADKTM 2894
QY 211 L-----R-----HC-----I-----R 216
Db 2895 INRGTSQSCMGGAIOQNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 2954
QY 217 -----EG--G-----Q-----D-----VP-----SNKD--VTSL--D- 232
Db 2955 PSCLDSEGHCEGTTKAPACQGGCVCEPDYVYLNKNCVPRIECCGCKDAQGV--LIPADK 3012
QY 233 -W-N-----SE-----G-----T--L-L--A----- 241
Db 3013 TWINRGTSQSCMGGAIOQNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3072
QY 242 -----T-----GS-----Y-----D-----G-F-AR--I----- 251
Db 3073 CQPSCNTDVHCEGSPNTLLSCREGCVQSGYVYLNKNCVPRIECCGCKDAQGVLIIPADK 3132
QY 252 -W-----T-----KDG-----N--L--A-S--T----- 261
Db 3133 TWITSGTQSCNCTGGAIQCNFQCPKTYCKDLKDGSSNCTNIPLQCPAHSHTNCLPTCQ 3192
QY 262 -----L-----G--Q-----HKGP-----IF-----AL-----K 273
Db 3193 CPLLCLDPGLCBGTSKVPSTCREGICQPGYVYLNKNCVPRIECCGCKDAQGVLIIPADK 3252
QY 274 -W-----N-K-----K-----G-----N-FI--L-- 282
Db 3253 TWISRGTSQSCMGGAIOQNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3312
QY 283 -----SA-----G-V--DK--T-----TII--W- 293
Db 3313 CVFSCNRCVTSFVSPSSCREGLCNHGFVFSDEKCVFRTQCGCKDAQGVLIIPADK 3372
QY 294 -----D-----AHT----- 297
Db 3373 SKGTQSCACVEGNIQCNFQCPKTYCKDLKDGSSNCTNIPLQCPAHSHTNCLPTCQ 3432
QY 298 -----G-----E-----A-----K--Q----- 302
Db 3433 SCLDPEGLCKDISPKVPSTCKEGCVQSGYVYLNKNCVPRIECCGCKDAQGVLIIPADK 3492
QY 303 -----Q-----F-----P--F-----HS----- 308
Db 3493 SPGCTQSCACMGGAIOQNFQCPKTYCKDLKDGSSNCTNIPLQCPAHSHTNCLPTCQ 3552
QY 309 -----A-----PA-----L-----D-----V--D--W- 316
Db 3553 SCLDPEGLCKGSPKVPSTCKEGICQSGYVYLNKNCVPRIECCGCKDAQGVLIIPADK 3612
QY 317 -----Q-----SN-----N-----T--F--A 323
Db 3613 SRGTQSCVCTGGSIOCLSSQCPGAYCKDNEDGNSNCTNIPLQCPAHSHTNCLPTCQ 3672
QY 324 SC-----ST-----D-----D----- 328
Db 3673 SCSDPEGHCEASGPRVLTSCREGCLNPGFVLDKCVPRVECCGCKDAQGVLIIPADK 3732
QY 329 -----M-----C--I-----H----- 332
Db 3733 SPGCTQSCACMGGAIOQNFQCPKTYCKDLKDGSSNCTNIPLQCPAHSHTNCLPTCQ 3792
QY 333 -----VCKL--Q-----D-----R-----P----- 341
Db 3793 SCLDPSALCGTSPKVPSTCKEGCVQSGYVYLNKNCVPRIECCGCKDAQGVLIIPADK 3852

QY 342 -----I-----K-T-----FQ--G-----H-T-----N 350
Db 3853 STGCIQSCACVEGNIQCNFQCPKTYCKDLKDGSSNCTNIPLQCPAHSHTNCLPTCQ 3912
QY 351 -----E-----V-----N-----A-----I--K-W--D- 358
Db 3913 LDGSCQTSKVPSTCKEGICQPGYVYLNKNCVPRIECCGCKDAQGVLIIPADK 3972
QY 359 -----P--T-----G--N-----L-----A- 365
Db 3973 TOSCACTGGAIOQNFQCPKTYCKDLKDGSSNCTNIPLQCPAHSHTNCLPTCQ 4032
QY 366 -----SCS--D-----SCS--D-----DM--TL-----K- 374
Db 4033 SHFTSCLPSCPPSCNLDGSCVESNFKAPSVCKGICQPGYVYLNKNCVPRIECCGCK 4092
QY 375 -----I-----W-----S-M-----K--Q----- 380
Db 4093 QGGLIPAGRTWISSDCTKSCMGIIQCRDQCPPTGYCKESNDSSRTCAKIPLQCPAH 4152
QY 381 -----D-----D-----N--CV--H--D- 386
Db 4153 SHYTNCPLPACRSCTDLGHCEGTSKVPSTCKEGICQPGYVYLNKNCVPRIECCGCK 4212
QY 387 -----L-----Q-----Q--HN-----K----- 392
Db 4213 QGGFVPAGRTWISRGTSKVPSTCKEGICQPGYVYLNKNCVPRIECCGCKDAQGV 4272
QY 393 -----E--I--Y-----T-----I 397
Db 4273 CLPSCLPSCDFPEGLCGGASPRAPSTCREGCVCEADYVLRDKCVLRTQCGCKDAQGV 4332
QY 398 -----K-----W-----S-----P----- 401
Db 4333 PANKTWLRGCAQCTCKGNIHCNFKPLTECKDSDVDGSSNCTKIPLQCPAHSHTY 4392
QY 402 -----T-----G-----P-----GT----- 406
Db 4393 CLPSCIPSCSNVDRCESTLQRPSTCEBGLCHSGFVFSDEKCVFRTQCGCKDAQGV 4452
QY 407 -----N-----P-----NAN-----L-M 414
Db 4453 PAGKWTGTSQRTCTGGLVQCHDFQCPGABQDIEDGNSNCTNIPLQCPAHSHTY 4512
QY 415 -L-----A-----S-----A----- 418
Db 4513 CLPQCPSCSDPDGHCESSTKAPACKGVCCEPDYVYLNKNCVPRIECCGCKDAQGV 4572
QY 419 -----S-----F----- 420
Db 4573 PESKTVWSRGCTKNCCTCKGTVQCHDFQCPGABQDIEDGNSNCTNIPLQCPAHSHTY 4632
QY 421 -----D-----D-----ST-----V-----R-----L- 426
Db 4633 CLPSCLPSCDFPEGLCGGTSPEVPSTCKEGICQSGYVYLNKNCVPRIECCGCKDAQGV 4692
QY 427 -----W----- 427
Db 4693 KTGQTVISSGCKICTCKGQFPQCPKTYCKDLKDGSSNCTNIPLQCPAHSHTY 4752
QY 428 -----D-----D-----D--V-----D-RG-- 432
Db 4753 CLPTCLPSCSNPDGRCEGTSKAPSTCREGCVQPGYVYLNKNCVPRIECCGCKDAQGV 4812
QY 433 -----I-----C--I-----H-----T----- 437
Db 4813 PAGNTWISSDCTQSCACTGVIQCNFVCPGSHQYNEGSSDCAANKLERCTIFGDPY 4872
QY 438 -LT-----K-----H----- 441
Db 4873 YLTFDGTFTYFLGRMNYLIKTVDKLPRGIEPLIMEGRNKISPKGSSHTLHEVTIVGYK 4932
QY 442 -----QE-----PVYS-----V--A-----F-----S-P- 452

```
Db 4933 IQQEBLVVNDKVAVP-YNENEHLRVMLRAQRLLLVTFEMVLDFOGKHSAVISLPT 4991
Qy 453 --DG--R-----Y-----LAS-----G-S-----F----- 462
Db 4992 TYRGLTRGLCGNYDRQSNELMLPSGVLTSNVHVFNGSNWEVKAQHAFPRPRALPEDEER 5051
Qy 463 DK-----C--V-----H----- 467
Db 5052 DEEPDLLQSECSQEQTALISSTQACRVLDVDPQGPFAACHQIIAPEPEQRCMLDMCTGWK 5111
Qy 468 -----I-----N-----NT--Q----- 472
Db 5112 TKEEBELRCRVLSGYAIIQOEAGANMTGWRDTHCAMTCPANTVYQRCMTPCPASCASFV 5171
Qy 473 --VC-----L--H-----Y-L-----N-----G 480
Db 5172 TPKVCEGPCVEGCASLPGYIYSDTQSLPVTHCGCTADGIYYKLGDSFVTNDCSQHCTCAS 5231
Qy 481 Q-VLLN--LG-R--SICL---YT-----L--P-H-----H----- 498
Db 5232 QGILLCEPYGCRAGES-CWVANETRCGFQDSPCLQNPCHNDGRCCEOGATFICHCDFGYG 5290
Qy 499 -----LVVI-P--LV-ALI-----E 509
Db 5291 GBFCTEPQDITTRKKIEASSLVAILPGLVLMVLPVLLPRVYVYMATRTTMGRRMRKKE 5350
Qy 510 --LL-----V-----LK 514
Db 5351 KXLLRQSLRLLEDADVPEPTFK 5372
```

Search completed: January 3, 2005, 15:28:12
Job time : 94.6667 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:55:11 ; Search time 102.333 Seconds
(without alignments)
2889.993 Million cell updates/sec

Title: US-09-987-701-4
Perfect score: 3649
Sequence: 1 MSISDEVNPLVRYLQESG.....LPHLVVIPLVALIELLVK 514

Scoring table: BLOSUM30
Gapop 1.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot_02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3437	94.2	514	1 TBLR HUMAN	Q9bk7 homo sapien
2	3405	93.3	514	1 TBLR MOUSE	Q8bhj5 mus musculus
3	3376.2	92.5	522	2 Q6GPC6	Q6gpc6 xenopus lae
4	3267.5	92.3	519	2 Q7SZM9	Q7szm9 xenopus lae
5	3216.1	88.1	527	1 TBLX MOUSE	Q9qxe7 mus musculus
6	3154.2	86.4	526	1 TBLX HUMAN	Q60907 homo sapien
7	3085.8	84.6	522	1 TBLY HUMAN	Q9bg87 homo sapien
8	2756.7	75.5	511	2 Q7Q371	Q7q371 anopheles g
9	2753.9	75.5	700	1 EBI DROME	Q95tj9 drosophila
10	2599.6	71.2	7073	2 AAP49011	Aap49011 sars coro
11	2598.3	71.2	20925	2 Q8NUX1	Q8nux1 trichoderma
12	2596.5	71.2	5636	2 Q6RCW7	Q96rw7 homo sapien
13	2592.4	71.0	34350	2 Q6XZ42	Q8wz42 homo sapien
14	2590.7	71.0	7191	2 Q6XA09	Q6xa09 alternaria
15	2590.7	71.0	7191	2 AAP78735	Aap78735 alternari
16	2590.6	71.0	7073	2 AAP82978	Aap82978 sars coro
17	2589.5	71.0	4599	1 LR1B MOUSE	Q9j118 mus musculus
18	2589.3	71.0	7962	2 Q10465	Q10465 homo sapien
19	2588.6	70.9	7073	1 R1AB CVHSA	P59641 h replicase
20	2588.6	70.9	7073	2 Q6JH48	Q6jh48 sars corona
21	2588.6	70.9	7073	2 Q6R7Y8	Q6r7y8 sars corona
22	2588.6	70.9	7073	2 Q6RCW7	Q6rcw7 sars corona
23	2588.6	70.9	7073	2 Q6RCX8	Q6rcx8 sars corona
24	2588.6	70.9	7073	2 Q6RCY9	Q6rcy9 sars corona
25	2588.6	70.9	7073	2 Q6RD00	Q6rd00 sars corona
26	2588.6	70.9	7073	2 Q6RD11	Q6rd11 sars corona
27	2588.6	70.9	7073	2 Q6RD22	Q6rd22 sars corona
28	2588.6	70.9	7073	2 Q6RD33	Q6rd33 sars corona
29	2588.6	70.9	7073	2 Q6RD44	Q6rd44 sars corona
30	2588.6	70.9	7073	2 Q6RD55	Q6rd55 sars corona
31	2588.6	70.9	7073	2 Q6RD66	Q6rd66 sars corona

RESULT 1

TBLR_HUMAN STANDARD; PRT; 514 AA.
AC Q9BK7; Q9H2I1, Q9H9A1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE F-box-like/WD-repeat protein TBLR1 (Nuclear receptor corepressor/HDAC3 complex subunit TBLR1) (TBL1-related protein 1).
DE complex subunit TBLR1; (TBL1-related protein 1).
GN Name=TBLR1; Synonyms=IRAL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR COMPLEX WITH NCOR1; NCOR2; GPS2; TBL1X AND HDAC3.
RX MEDLINE=21930350; PubMed=11931768;
RT Zhang J., Kalkum M., Chait B.T., Roeder R.G.;
RT "The N-Cor-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";
RL Mol. Cell 9:611-623(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=11063877;
RT Zhang X., Dormady S.P., Basch R.S.;
RT "Identification of four human CDNAs that are differentially expressed by early hematopoietic progenitors.";
RL Exp. Hematol. 28:1286-1296(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ngl1285;
RT Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahata K., Murakami K., Yasuda T., Iwawana T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa K., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hirakawa S., Chiba Y., Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S., Yoshioka K., Matsunawa H., Ichihara T., Shiohara H., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

ALIGNMENTS

32 2588.6 70.9 7073 2 Q6TPE9
33 2588.6 70.9 7073 2 Q6UZF1
34 2588.6 70.9 7073 2 Q6UZF5
35 2588.6 70.9 7073 2 Q6V586
36 2588.6 70.9 7073 2 Q6VA80
37 2588.6 70.9 7073 2 Q6VA91
38 2588.6 70.9 7073 2 Q6VAA2
39 2588.6 70.9 7073 2 AAR14802
40 2588.6 70.9 7073 2 AAR14806
41 2588.6 70.9 7073 2 AAR14810
42 2588.6 70.9 7073 2 AAR91584
43 2588.6 70.9 7073 2 AAS00002
44 2588.6 70.9 7073 2 AAP94735
45 2588.6 70.9 7073 2 AAP94746

Q6tpe9 sars corona
Q6uzf1 sars corona
Q6uzf5 sars corona
Q6v586 sars corona
Q6v80 sars corona
Q6va91 sars corona
Q6vaa2 sars corona
Aar14802 sars coro
Aar14806 sars coro
Aar14810 sars coro
Aar91584 sars coro
Aas00002 sars coro
Aap94735 sars coro
Aap94746 sars coro

RA Okitani R., Kawakami T., Nouguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama H., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RT Nat. Genet. 36:40-45 (2004).
RN [4]
RN SEQUENCE OF 1-125 FROM N.A.
RP TISSUE=Adrenal cortex;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallaion D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [5]
RN COMPONENT OF THE N-COR COMPLEX WITH TBL1X; CORO2A AND HDAC3, AND
RP HISTONE-BINDING.
RX PubMed=12628926;
RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,
RA Wong J.;
RT "Purification and functional characterization of the human N-Cor
RT complex: the roles of HDAC3, TBL1 and TBL1X.";
RL EMBO J. 22:1336-1346 (2003).
RN [6]
RN FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.
RX PubMed=14980219;
RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;
RT "A corepressor/coactivator exchange complex required for
RT transcriptional activation by nuclear receptors and other regulated
RT transcription factors.";
RL Cell 116:511-526 (2004).
CC -1- FUNCTION: F-box-like protein involved in the recruitment of the
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
CC transcription units. Plays an essential role in transcription
CC activation mediated by nuclear receptors. Probably acts as
CC integral component of the N-Cor corepressor complex that mediates
CC the recruitment of the 19S proteasome complex, leading to the
CC subsequent proteasomal degradation of N-Cor complex, thereby
CC allowing cofactor exchange, and transcription activation.
CC -1- SUBUNIT: Component of the N-Cor repressor complex, at least
CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.
CC Probable component of some E3 ubiquitin ligase complex. Interacts
CC with histones H2B and H4.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- DOMAIN: The F-box-like domain is related to the F-box domain, and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes (By similarity).
CC -1- SIMILARITY: Belongs to the WD-repeat E3 family.
CC -1- SIMILARITY: Contains 1 F-box-like domain.
CC -1- SIMILARITY: Contains 1 Lish domain.
CC -1- SIMILARITY: Contains 8 WD repeats.
CC -1- CAUTION: Ref.4 sequence differs from that shown due to a potential
CC poly-A sequence starting in position 125.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF314544; AAK00301.1; -;
DR EMBL; AF268193; AAG44736.1; -;
DR EMBL; AK022956; BAB14331.1; -;
DR EMBL; BC060320; AAH60320.1; ALT_TERM.
DR HSSP; P16649; 1ERJ.
DR MIM; 608628; -;
DR InterPro; IPR006594; Lish.
DR InterPro; IPR011047; Quin_alc_DH_like.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Activator; Nuclear protein; Repeat; Transcription regulation;
KW Ub1 conjugation pathway; WD repeat.
FT DOMAIN 4 36
FT Lish.
FT DOMAIN 41 86
FT F-box-like.
FT REPEAT 167 206
FT WD 1.
FT REPEAT 223 282
FT WD 2.
FT REPEAT 264 303
FT WD 3.
FT REPEAT 306 344
FT WD 4.
FT REPEAT 347 386
FT WD 5.
FT REPEAT 389 437
FT WD 6.
FT REPEAT 440 479
FT WD 7.
FT REPEAT 481 513
FT WD 8.
FT DOMAIN 108 118
FT Poly-Ala.
FT CONFLICT 31 31
FT E -> K (in Ref. 2).
FT CONFLICT 59 59
FT Y -> H (in Ref. 3).
FT CONFLICT 389 389
FT A -> Q (in Ref. 2).
SQ SEQUENCE 514 AA; 55595 MW; 0B556D2EE4BA796D CRC64;
Query Match 94.2%; Score 3437; DB 1; Length 514;
Best Local Similarity 90.4%; Pred. No. 5.7e-62;
Matches 483; Conservative 6; Mismatches 5; Indels 40; Gaps 12;
QY 1 MTSIDVNFVLYRYLOESGFSHSAFTFGIKSHISOSNINGALVPPAALISIIQKLOYV 60
DB 1 MTSIDVNFVLYRYLOESGFSHSAFTFGIESHISOSNINGALVPPAALISIIQKLOYV 60
QY 61 EAEVSINEDGTLFDGRIEISLSLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQ 120
DB 61 EAEVSINEDGTLFDGRIEISLSLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQ 120
QY 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGDIPIPNKAVLRGHESEVPICAWNPV 180
DB 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGDIPIPNKAVLRGHESEVPICAWNPV 180
QY 181 SLLASGSGDSTARIWNLSNSTSGSTQLVRHCITREGQDVPNSKNDVTSLDWNSGTL 240
DB 181 SLLASGSGDSTARIWNLSNSTSGSTQLVRHCITREGQDVPNSKNDVTSLDWNSGTL 240
QY 241 ATGSDYGFARIWTKDGNLSTLQHKGPFIKLNKKNKGNFILSAGVDKTTIIWDAHTGEA 300
DB 241 ATGSDYGFARIWTKDGNLSTLQHKGPFIKLNKKNKGNFILSAGVDKTTIIWDAHTGEA 300
QY 301 KQOFFPHSPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNTNEVNAIKWDPT 360
DB 301 KQOFFPHSPALDWDVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNTNEVNAIKWDPT 360

```
OY 361 GNLLASCSDDMTLKIMSKQDNVCVHDLQOHKEIYITIKWSPGPGTNNPNANMLASASF 420
Db 361 GNLLASCSDDMTLKIMSKQDNVCVHDLQAHKEIYITIKWSPGPGTNNPNANMLASASF 420
OY 421 DSTVRLWVDVDRGICHTLTKHQPVPVSVAFSPDGRVLAGSDPKCVHINTQ-----VCLH 476
Db 421 DSTVRLWVDVDRGICHTLTKHQPVPVSVAFSPDGRVLAGSDPKCVHINTQGTALV--H 478
OY 477 -Y--LNG--QVLIN-----LGR-----SICLYTLPHLHVLPVLIALLVL-K 514
Db 479 SYRGTGIEVCWNAAGDKVGASDGSVC-----V-L-D-----LRK 514

RESULT 2
TBLR MOUSE
ID TBLR MOUSE STANDARD; PRT; 514 AA.
AC QBHJ75; Q8CEG4; Q8VEG3; Q9EQD4;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE F-box-like/WD-repeat protein TBLR1 (Nuclear receptor corepressor/HDAC3
DE complex subunit TBLR1) (TBLR1-related protein 1).
GN Name=Tblr1; Synonyms=Iral;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Zhang X., Dormady S., Basch R.;
RT "Identification of four human cDNAs that are differentially expressed
RT by early hematopoietic progenitors.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]_
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Testis;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Oosato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanpin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perteau G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sadelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verdano R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski B., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [3]_
RP SEQUENCE OF 314-514 FROM N.A.
RC TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

```
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
CC transcription units. Plays an essential role in transcription
CC activation mediated by nuclear receptors. Probably acts as
CC integral component of the N-CoR corepressor complex that mediates
CC the recruitment of the 19S proteasome complex, leading to the
CC subsequent proteosomal degradation of N-CoR complex, thereby
CC allowing cofactor exchange, and transcription activation (By
CC similarity).
CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least
CC composed of NCOR1, NCOR2, HDAC3, TBLR1, TBLR2, and GPS2.
CC Probable component of some E3 ubiquitin ligase complex. Interacts
CC with histones H2B and H4 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes (By similarity).
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -!- SIMILARITY: Contains 1 F-box-like domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF268195; AAG44738.1; -
DR EMBL; AK029595; BAC26526.1; -
DR EMBL; AK033347; BAC28241.1; -
DR EMBL; AK036064; BAC29294.1; -
DR EMBL; BC018512; AAH18512.1; -
DR HSSP; P16649; 1ERJ.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR011047; Quin_alc_DH_like.
DR InterPro; IPR001580; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00882; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Activator; Nuclear protein; Repeat; Transcription regulation;
KW Ubiquitin conjugation pathway; WD repeat.
FT DOMAIN 4 36 Lish.
FT DOMAIN 41 86 F-box-like.
FT REPEAT 167 206 WD 1.
FT REPEAT 223 262 WD 2.
FT REPEAT 264 303 WD 3.
```

```

FT REPEAT 306 344 WD 4.
FT REPEAT 347 386 WD 5.
FT REPEAT 389 437 WD 6.
FT REPEAT 440 479 WD 7.
FT REPEAT 481 513 WD 8.
FT DOMAIN 108 117 Poly-Ala.
FT CONFLICT 185 185 A -> V (in Ref. 1).
FT CONFLICT 374 374 K -> R (in Ref. 2; BAC29294).
FT CONFLICT 480 480 Y -> C (in Ref. 3).
SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;

Query Match 93.3%; Score 3405; DB 1; Length 514;
Best Local Similarity 89.5%; Pred. No. 2.9e-61;
Matches 478; Conservative 7; Mismatches 9; Indels 40; Gaps 12;

QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVSVINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAAASQ 120
Db 61 EAEVSVINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAAATNQ 120

QY 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWN 180
Db 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWN 180

QY 181 SDLLASGSDSTARIWNLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
Db 181 SDLLASGSDSTARIWNLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240

QY 241 ATGSDGFAFIWTKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTIIWDAHTGEA 300
Db 241 ATGSDGFAFIWTKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTIIWDAHTGEA 300

QY 301 KQFPFHSAPALDVQSNNTFASCSTDMCIHVCKLQDQRPDKTFOGHTNEVNAIKWDPT 360
Db 301 KQFPFHSAPALDVQSNNTFASCSTDMCIHVCKLQDQRPDKTFOGHTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKIMSKQDCNVHDLQHNKEIYIKSPGPGTNNPNANMLASAF 420
Db 361 GNLLASCSDDMTLKIMSKQDCNVHDLQHNKEIYIKSPGPGTNNPNANMLASAF 420

QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFSPGRIYASGPKCVHIWNTQ----VCLH 476
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFSPGRIYASGPKCVHIWNTQ----VCLH 476

QY 477 -Y--LNG-----QVLLN-----LGR-----SICLYTLPHLHVVIPLVLIBLVL-K 514
Db 477 SYRGTGIGFVVCWNAAGDKVGSASDGSVC-----V-L-D-----LRK 514

```

RESULT 3

```

Q6GPC6 PRELIMINARY; PRT; 522 AA.
ID Q6GPC6
AC Q6GPC6;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

```

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.R., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Contains 8 WD repeats.
DR EMBL: BC073215; AAH73215.1; -.
DR InterPro: IPR006594; Lish.
DR InterPro: IPR011047; Quin_alc_DH_like.
DR InterPro: IPR011680; WD40.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00320; WD40; 8.
DR PROSITE: PS00896; Lish; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 4.
DR PROSITE: PS00882; WD_REPEATS_2; 6.
DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 522 AA; 56284 MW; 4FB0F58F503B461F CRC64;

```

Query Match 92.5%; Score 3376.2; DB 2; Length 522;
Best Local Similarity 87.5%; Pred. No. 1.2e-60;
Matches 474; Conservative 9; Mismatches 11; Indels 48; Gaps 15;

```

QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVSVINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQO-----AAAAA 116
Db 61 EAEVSVINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAA 120

QY 117 AA-S---QQGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEV 172
Db 121 AATSANNQPPAKNGESTANGENGHALANNHTDMMEVDGVEIPPNKAVVLRGHESEV 180

QY 173 FICAWNPSDILLASGSDSTARIWNLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLD 232
Db 181 FICAWNPSDILLASGSDSTARIWNLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLD 240

QY 233 WNSEGTLATGSDYDGFARITWKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTII 292
Db 241 WNSEGTLATGSDYDGFARITWKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTII 300

QY 293 WDAHTGEAKQQPPFHSAPALDVQSNNTFASCSTDMCIHVCKLQDQRPDKTFOGHTNEV 352

```



```

Db 301 WDAHTGAQKQPPHSPALDWDQSNNTFASCSMDCHVCKLQGDPRPKTFQGHTEV 360
Qy 353 NAIKWDPTGNLLASCSDMTLKIWSMKQDNCVHDLQOHKEIYTIKWSPTGPTGNPNAN 412
Db 361 NAIKWDPTGNLLASCSDMTLKIWSMKHDTCVHDLQAHKEIYTIKWSPTGPTGNPNAN 420
Qy 413 LMLASAFDSTVRLMDVDRGICHTLTKHQPVPYVAFSPDGRYLASGDFKCVHIWNTQ 472
Db 421 LMLASAFDSTVRLMDVDRGICHTLTKHQPVPYVAFSPDGRYLASGDFKCVHIWNTQ 480
Qy 473 -----VCLH-Y--LNG--QVLLN-----LGR-----SICLYTLPHHLVWIPVLAIELLVL 513
Db 481 TGAIV--HSYRTGGIGFVCWNAAGDKVGASASDGSVC-----V-L-D---L 520
Qy 514 -K 514
Db 521 RK 522

RESULT 4
ID Q7SZM9 PRELIMINARY; PRT; 519 AA.
AC Q7SZM9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12794076;
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.;
RT "Fusion Protein of Retinoic Acid Receptor [alpha] with Promyelocytic
RT Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein
RT Recruits N-CoR-TBLR1 Corepressor Complex to Repress Transcription in
RT Vivo.";
RL J. Biol. Chem. 278:30788-30795 (2003).
CC -I- SIMILARITY: Contains 8 WD repeats.
DR EMBL; AY225088; AAP20646.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Receptor; Repeat; WD repeat.
SQ SEQUENCE 519 AA; 56043 MW; 5E998EDC8C92296 CRC64;

Query Match 92.3%; Score 3367.5; DB 2; Length 519;
Best Local Similarity 87.8%; Pred. No. 1.9e-60;
Matches 472; Conservative 11; Mismatches 11; Indels 45; Gaps 14;

Qy 1 MSISDEVNPLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIKGLQVY 60
Db 1 MSISDEVNPLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIKGLQVY 60
Qy 61 EAFVSINEDGTLFDGRPIESLSLIDVMPDVVQTRQAYRDKLAQQQ--AAAAA 118
Db 61 EAEVSINEDGTLFDGRPIESLSLIDVMPDVVQTRQAYRDKLAQQQTRAAAAA 120
Qy 119 S---QQGSAKNGENTANGENGATHTANNHTDMMEVDGDVEIPPNKAVVLRGHSEVFIC 175
Db 121 TPNNOQPPAKNGENTANGENGHALNHNHTDMMEVDGDVEIPSPKAVVLRGHSEVFIC 180

```

RESULT 5

TBLX_MOUSE

```

ID TBLX_MOUSE STANDARD; PRT; 527 AA.
AC Q9QXE7; Q8BMM0; Q8BYQ4; Q8COAL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/wd-repeat protein TBLX (Traneducin beta-like 1X protein).
GN Name=Tblx; Synonyms=Tbll;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC TISSUE=Hypothalamus, Medulla oblongata, and Pituitary;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Oosato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.W.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Kongsaga A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltasa L., Marchionni L., Mckenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi F., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Savelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takemura Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski B., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [2]

```


Db 523 -L-D---LRK 527

RESULT 6

TBLX_HUMAN

ID TBLX_HUMAN STANDARD; PRT; 526 AA.

AC O60907; Q86UY2; 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE F-box-like/WD-repeat protein TBL1X (Transducin beta-like IX protein)

DE (Transducin-beta-like 1, X-linked) (SMP55).

GN Name=TBL1X; Synonyms=TBL1;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DISEASE.

RX MEDLINE=99264241; PubMed=10330347;

RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,

RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.,

RT "X-linked late-onset sensorineural deafness caused by a deletion

RT involving OAI and a novel gene containing WD-40 repeats.";

RL Am. J. Hum. Genet. 64:1604-1616(1999).

RN [2]

RP SEQUENCE FROM N.A.

RX TISSUE=Lymph, and Pancreas;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klautner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [3]

RP IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR

RP COMPLEX WITH NCOR2 AND HDAC3.

RX PubMed=10809664;

RA Guenther M.G., Lane W.S., Fischle W., Verdin E., Lazar M.A.,

RA Shiekhattar R.;

RT "A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-

RT repeat protein linked to deafness.";

RL Genes Dev. 14:1048-1057(2000).

RN [4]

RP COMPONENT OF THE N-COR COMPLEX WITH NCOR2 AND HDAC3.

RX PubMed=10944117;

RA Li J., Wang J., Wang J., Nawaz Z., Liu J.M., Qin J., Wong J.;

RT "Both corepressor proteins SMRT and N-COR exist in large protein

RT complexes containing HDAC3.";

RL EMBO J. 19:4342-4350(2000).

RN [5]

RP SUBUNIT OF A COMPLEX WITH UBE2D1, CACYBP, SIAH1 AND APC.

RX PubMed=11389839;

RA Matsuzawa S.-I., Reed J.C.;

RT "Siah-1, SIP, and Ebi collaborate in a novel pathway for beta-catenin

RT degradation linked to p53 responses.";

RL Mol. Cell 7:915-926(2001).

RN [6]

RP COMPONENT OF THE N-COR COMPLEX WITH NCOR1, NCOR2, GPS2, TBL1R AND

RP HDAC3.

RX MEDLINE=21930350; PubMed=11931768;

RA Zhang J., Kalkum M., Chait B.T., Roeder R.G.;

RT "The N-COR-HDAC3 nuclear receptor corepressor complex inhibits the JNK

RT pathway through the integral subunit GPS2.";

RL Mol. Cell 9:611-623(2002).

RN [7]

RP COMPONENT OF THE N-COR COMPLEX WITH TBL1R, CORO2A AND HDAC3, AND

RP HISTONE-BINDING

RX PubMed=12628926;

RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,

RA Wong J.;

RT "Purification and functional characterization of the human N-COR

RT complex: the roles of HDAC3, TBL1 and TBL1R.";

RL EMBO J. 22:1336-1346(2003).

RN [8]

RP FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.

RX PubMed=14980219;

RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;

RT "A corepressor/coactivator exchange complex required for

RT transcriptional activation by nuclear receptors and other regulated

RT transcription factors.";

RL Cell 116:511-526(2004).

CC -!- FUNCTION: F-box-like protein involved in the recruitment of the

CC ubiquitin/19S proteasome complex to nuclear receptor-regulated

CC transcription units. Plays an essential role in transcription

CC activation mediated by nuclear receptors. Probably acts as

CC integral component of corepressor complexes that mediates the

CC recruitment of the 19S proteasome complex, leading to the

CC subsequent proteasomal degradation of transcription repressor

CC complexes, thereby allowing cofactor exchange.

CC -!- SUBUNIT: Component of the N-COR repressor complex, at least

CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.

CC Component of a E3 ubiquitin ligase complex containing UBE2D1,

CC SIAH1, CACYBP/SIP, SKP1A, APC and TBL1X. Probably part of other

CC corepressor complexes, that do not contain NCOR1 and NCOR2.

CC Interacts with histones H2B, H3a and H4.

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -!- TISSUE SPECIFICITY: Ubiquitous.

CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and

CC apparently displays the same function as component of ubiquitin E3

CC ligase complexes (By similarity).

CC -!- DISEASE: Defects in TBL1X may be a cause of late-onset

CC sensorineural deafness (OASD) (MIM:300650). OASD is an X-linked

CC recessive disorder characterized by ocular albinism and

CC progressive sensorineural hearing loss in the fourth and fifth

CC decades of life. OASD may be caused by deletion of both GPR143/OAI

CC and TBL1X adjacent genes; TBL1X defects possibly causing the

CC hearing phenotype.

CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.

CC -!- SIMILARITY: Contains 1 F-box-like domain.

CC -!- SIMILARITY: Contains 1 Lish domain.

CC -!- SIMILARITY: Contains 8 WD repeats.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL; Y12781; CAAY73319.1; ALT_INIT.

DR EMBL; BC032708; AAH32708.1; -.

DR EMBL; BC052304; AAH52304.1; ALT_INIT.

DR HSSP; P16649; 1ERJ.

DR Genew; HGNC:11585; TBL1X.

DR MIM; 300196; -.

DR MIM; 300650; -.

DR GO; GO:0007605; P:perception of sound; TAS.

DR GO; GO:0007165; P:signal transduction; TAS.

DR GO; GO:0007601; P:visual perception; TAS.

```

DR InterPro: IPR006594; Lish.
DR InterPro: IPR011047; Quin_alc_DH_like.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00320; WD40; 8.
DR PROSITE: PS00896; LISH; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 4.
DR PROSITE: PS00082; WD_REPEATS_2; 6.
DR PROSITE: PS00294; WD_REPEATS_REGION; 1.
KW Activator; Deafness; Nuclear protein; Repeat;
KW Transcription regulation; Ub1 conjugation pathway; WD repeat.
FT DOMAIN 4 36 Lish.
FT REPEAT 41 86 F-box-like.
FT REPEAT 179 218 WD 1.
FT REPEAT 235 274 WD 2.
FT REPEAT 276 315 WD 3.
FT REPEAT 318 358 WD 4.
FT REPEAT 359 398 WD 5.
FT REPEAT 401 449 WD 6.
FT REPEAT 452 491 WD 7.
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 Poly-Ala.
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F659 CRC64;

Query Match      86.4%; Score 3154.2; DB 1; Length 526;
Best Local Similarity 78.3%; Pred. No. 8e-56;
Matches 430; Conservative 41; Mismatches 20; Indels 58; Gaps 22;

Qy 1 MSISDEVNVLVRYLQESGFSHSATFGIKSHISOSNINGALVPPAALISIIKGLQYV 60
Db 1 MSISDEVNVLVRYLQESGFSHSATFGIESHSOSNINGTLVPPAALISIIKGLQYV 60

Qy 61 EAEVISEDGTLFDGRPIESLSLIDAMPDVVQTRQAVRDKLAQQ--AAAAAAA-AA 117
Db 61 EAEVISEDGTVDFGRPIESLSLIDAMPDVVQTRQAFREKLAQQASAAAAATAA 120

Qy 118 ASQ-----QG-S-----AKNGENTANGENGHAHTIANHTDM-MEVDGDVEIPPNKAVLR 166
Db 121 ATAAATTSAGVSHQNFPSKNREATVNGEENRAHSV-NNHA-KPMEIDGEVEIPSSKATVLR 178

Qy 167 GHESEVFIICAWNPSVLLASGSGDSTARIWNLSNSTSGSTQLVLRHCIREGGDVPSPNK 226
Db 179 GHESEVFIICAWNPSVLLASGSGDSTARIWNLSNENSGSTQLVLRHCIREGGHDVPSPNK 238

Qy 227 DVTSLDMNSGTLTATGSDYDGFARIWTKDGNLSTLQHKGPFPALKWKKGNFILLSAGV 286
Db 239 DVTSLDMNTGTLTATGSDYDGFARIWTEGDNLSTLQHKGPFPALKWKNRKNVILLSAGV 298

Qy 287 DKTIIWDATGAKQOPPHSAPALDVQSN--TFASCTDMCHVCKLGQDRIPKTF 345
Db 299 DKTIIWDATGAKQOPPHSAPALDVQ--NNTTFASCTDMCHVCKLGCDRVPKTF 357

Qy 346 QGHTNEVNAIKWPTGNLASCDDMTLTKIWSKQDNCVHDLOQHNKEIYTIKWSPTGPG 405
Db 358 QGHTNEVNAIKWPTGNLASCDDMTLTKIWSKQEVCHIDLQAHNKEIYTIKWSPTGPA 417

Qy 406 TNNPNANMLASFSFSTVRLWDVDRGICHTLTGKQEPVYSVAFSPDGRYLASGFDKC 465
Db 418 TSNPNSNIMLASFSFSTVRLWDIERGVCTHTLTGKQEPVYSVAFSPDGRYLASGFDKC 477

Qy 466 VHTWNTQ-----VCLH-Y--LNG--QVLLN-----LGR-----SICLYTLPHLWVPLVA 506
Db 478 VHTWNTQSNLV--HSYRGTTGGIFECVWNRAGDKVGASASDGSVC-----V-- 521

Qy 507 LIELLVL-K 514
Db 522 L-D---LRK 526

RESULT 7

```

```

TBLY_HUMAN
ID TBLY_HUMAN STANDARD; PRT; 522 AA.
AC Q9BQ87;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein TBLLY (Transducin beta-like 1V protein)
DE (Transducin-beta-like 1, Y-linked).
GN Name=TBLLY; Synonyms=TBLL1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND TISSUE SPECIFICITY
RX MEDLINE=22699787; PubMed=12815422; DOI=10.1038/nature01722;
RA Skaletsky H., Kuroda-Kawaguchi T., Minx P.J., Cordum H.S.,
RA Hillier L.W., Brown L.G., Repping S., Pyntikova T., Ali J., Bieri T.,
RA Chinwalla A., Delehaanty A., Delehaanty K., Du H., Fewell G.,
RA Fulton L., Fulton R., Graves T.A., Hou S.-F., Latreille P.,
RA Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W.,
RA Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K.,
RA Schultz B., Strong C., Tin-Wollam A., Yang S.-P., Waterston R.H.,
RA Wilson R.K., Rozen S., Page D.C.;
RT "The male-specific region of the human Y chromosome is a mosaic of
RT discrete sequence classes.";
RL Nature 423:825-837(2003).
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
CC transcription units. Plays an essential role in transcription
CC activation mediated by nuclear receptors. Probably acts as
CC integral component of corepressor complexes that mediates the
CC recruitment of the 19S proteasome complex, leading to the
CC subsequent proteosomal degradation of transcription repressor
CC complexes, thereby allowing cofactor exchange (By similarity).
CC -!- SUBUNIT: Probable component of the N-CoR repressor complex and
CC some E3 ubiquitin ligase complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: Fetal brain and prostate.
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes (By similarity).
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -!- SIMILARITY: Contains 1 F-box-like domain.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF332220; AAK13472.1; -.
DR EMBL: AF332221; AAK13473.1; -.
DR EMBL: AF332222; AAK13474.1; -.
DR HSSP: P16649; IERJ.
DR Genew; HGNC:18502; TBLLY.
DR MIM; 400033; -.
DR InterPro: IPR006594; Lish.
DR InterPro: IPR001680; WD40.
DR InterPro: IPR011046; WD40_like.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00320; WD40; 8.
DR PROSITE: PS00896; LISH; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 4.
DR PROSITE: PS00082; WD_REPEATS_2; 6.
DR PROSITE: PS00294; WD_REPEATS_REGION; 1.

```

KW Activator; Nuclear protein; Repeat; Transcription regulation;
 KW Ubl conjugation pathway; WD repeat.
 FT DOMAIN 4 36 Lish.
 FT DOMAIN 41 86 F-box-like.
 FT REPEAT 177 216 WD 1.
 FT REPEAT 233 272 WD 2.
 FT REPEAT 274 313 WD 3.
 FT REPEAT 316 354 WD 4.
 FT REPEAT 357 396 WD 5.
 FT REPEAT 399 447 WD 6.
 FT REPEAT 450 489 WD 7.
 FT REPEAT 491 521 WD 8.
 SQ SEQUENCE 522 AA; 5668 MW; 4E020216422442D8 CRC64;

Query Match 84.6%; Score 3085.8; DB 1; Length 522;
 Best Local Similarity 78.0%; Pred. No. 2.4e-54;
 Matches 423; Conservative 37; Mismatches 30; Indels 52; Gaps 22;

QY 1 MSISDEVNPLVRYLQESGFSAFTFGIKSHISOSNINGALVPPAALISIIQKGLQYV 60
 DB 1 MSITSDEVNPLVRYLQESGFSAFTFGIESHISOSNINGTLVPPSALISIIQKGLQYV 60

QY 61 EAEVSINEDGTLPDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQA-AAA--A-A-114
 DB 61 EAEISINKDGTVPDSRPIESLSLIVAVIPDVQMROQAFGEKLTQQQASAAATEASAMAK 120

QY 115 AA-----AA-SQGSACKNGENTANGENGAHTIANNHTDM-MEVDGVEIPPNKAVVLRGH 168
 DB 121 AATWTPAAISQQNPPKRRATVNGEAGAEI-NNHS-KPMEIDGVEIPPNKATVLRGH 178

QY 169 ESEVFTICANNPVSDDLASGSGDSTARIMNLSNSTSGSTQLVLRHCIRREGQDVPNSKDV 228
 DB 179 ESEVFTICANNPVSDDLASGSGDSTARIMNLSNSNGSTQLVLRHCIRREGHDPNSKDV 238

QY 229 TSLDWNSEGTILATGSDGFARITWTKDGNLSTLQGHKGPFPALKKWKNGNFIISAGVDK 288
 DB 239 TSLDWNSEGTLLAMGSDGFARITWENGNLSTLQGHKGPFPALKKWKNGNYLSAGVDK 298

QY 289 TTIINDAHTGEAKQPPFHSAPALPDVDWQSN--TPASCSTDMCIHVCKLGQDRPIKTFQ 347
 DB 299 TTIINDAHTGEAKQPPFHSAPALPDVDWQ--NNMTFASCSTDMCIHVCRGCDHPVKTFQ 357

QY 348 HTNEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGPGTN 407
 DB 358 HTNEVNAIKWDPSGMLASCDDMTLKIWSMKQDACVHDLQHSKEIYTIKWSPTGPTS 417

QY 408 NPVNAIMLASAFDSTVRLVDVDRGICHTLTKHQBPVYSVAFSPDGRYLASGSPDKCVH 467
 DB 418 NPNSSIMLASAFDSTVRLVDVEQGVCTHTLMKHQBPVYSVAFSPDGKYLASGSPDKYVH 477

QY 468 IWNTO-----VCLH-Y--LNG--QVLLN-----LGR-----SICLYTLPHLHVPIPLVALI 508
 DB 478 IWNTOGSLV--HSYQGTGGIPEVCNARGDKVGASASDGSVC-----V-L-520

QY 509 EL 510
 DB 521 DL 522

RESULT 8
 QY0371 PRELIMINARY; PRT; 511 AA.
 AC QY0371;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE AgCp11563 (Fragment).
 GN Name=agc45469; ORFNames=ENSANG00000019755;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -|- SIMILARITY: Contains 8 WD repeats.
 CC -|- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008964; EAA12470.1; --
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF04000; WD40; 8.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000018; WD40; 3.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 3.
 DR PROSITE; PS00882; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT NON TER 1
 SQ SEQUENCE 511 AA; 55631 MW; 7C4C2670316A8580 CRC64;

Query Match 75.5%; Score 2756.7; DB 2; Length 511;
 Best Local Similarity 68.4%; Pred. No. 3.1e-47;
 Matches 379; Conservative 49; Mismatches 33; Indels 93; Gaps 39;

QY 1 MSISDEVNPLVRYLQESGFSAFTFGIKSHISOSNINGALVPPAALISIIQKGLQYV 60
 DB 8 MFSSDEVNPLVRYLQESGFSAFTFGIESHISOSNINGALVPPAALISIIQKGLQYV 67

QY 61 EAEVSINEDGT---LFDGRPIESLSLIDAVMPDVVQTRQ--QAYRDKLAQQ-QAAAAA 114
 DB 68 EAEISIGEDGTQRL--G---ESLSLIDAVMEVVARNQMQ-----NQKQ----- 109

QY 115 AAAASQGSACKNGENTANG--BENGHAH-----TIANNHT-----DMVEVDGVEIPPNKA 162
 DB 110 -ATKTEP---P--E-T-NCTTEE-PAPPPPTTTTATT-TPAPVETMEVDQSIETPASKA 159

QY 163 VVLRGHESEVFTICANNPVS--DLLASGSGDSTARIMNLSNSTSGSTQLVLRHCIRREGQD 221
 DB 160 TVLRGHESEVFTICANNP--STDLLASGSGDSTARIMNLSNSTSGSTQLVLRHCIRREGQD 217

QY 222 VPSNKKDVTSLDWNSEGTILATGSDGFARITW--TKDGNLSTLQGHKGPFPALKKWKNGNF 280
 DB 218 VPSNKKDVTSLDWNSEGTILATGSDGFARITW--TKDGNLSTLQGHKGPFPALKKWKNGNY 276

QY 281 ILSAGVDKTTIINDAHTGE--AKQPPFHSAPALPDVDWQSNNTFASCSTDMCIHVCKLGQD 339
 DB 277 ILSAGVDKTTIINDAHTGE--AKQPPFHSAPALPDVDWQSNNTFASCSTDMCIHVCKLGQD 335

QY 340 RIKTFOGHTNEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHDLQHNKEIYTIKW 399
 DB 336 KPIKFOGHTNEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHDLQHNKEIYTIKW 395

QY 400 SPTGPGTNPNANMLASAFDSTVRLVDVDRGICHTLTKHQBPVYSVAFSPDGRYLAS 459
 DB 396 SPTGPGTNPNANMLASAFDSTVRLVDVDRGICHTLTKHQBPVYSVAFSPDGRYLAS 455

QY 460 GSFDKCVHIWNTQ-----VCLH-YLNGQVLLNLSRST---CLY-----TLPHHLV----- 500
 DB 456 GSFDKCVHIWNTQSGQLV--HSY-KG-----TG-GIFEVC-WNSRGS-K---VGASASDG 501

QY 501 -VPLVALIEL--L 511
 DB 502 SV--FV-L-DLRKL 511

RESULT 9
 EBI_DROME
 ID EBI_DROME STANDARD; PRT; 700 AA.
 AC Q95R39; Q9XZK1;
 DT 05-JUL-2004 (Rel. 44, Created)
 DT 05-JUL-2004 (Rel. 44, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein ebi.
GN Name-ebi; ORFNames-CG4063;
OS Drosophila melanogaster. [Fruit fly].
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND MUTANT EBI4.
RX MEDLINE=99234084; PubMed=10215623;
RA Dong X., Tsuda L., Zavitz K.H., Lin M., Li S., Carthew R.W.,
RA Zipursky S.L.;
RT "ebi regulates epidermal growth factor receptor signaling pathways in
RT Drosophila.";
RL Genes Dev. 13:954-965 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RP STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
RN [4]
RP FUNCTION, INTERACTIONS WITH PHYL AND SINA, AND MUTANTS EBI1C1 AND
RP EBI1C3.
RX PubMed=11032805;
RA Boulton S.J., Brook A., Staehling-Hampton K., Heitzler P., Dyson N.;
RT "A role for Ebi in neuronal cell cycle control.";
RL EMBO J. 19:5376-5386 (2000).
RN [5]
RP FUNCTION, AND INTERACTIONS WITH SNO; SMR AND SU(H).
RP PubMed=12230979;
RA Tsuda L., Nagaraj R., Zipursky S.L., Banerjee U.;
RT "An EGFR/Ebi/Sno pathway promotes delta expression by inactivating
RT Su(H)/SMRTER repression during inductive notch signaling.";
RL Cell 110:625-637 (2002).
CC [1-] FUNCTION: F-box-like component of E3 ubiquitin ligase complexes;
CC involved in R7 photoreceptor cell differentiation, cone cell
CC development and neuronal cell cycle control. E3 ubiquitin ligase
CC complexes mediate ubiquitination and subsequent proteasomal
CC degradation of target proteins. Required for specification of R7
CC photoreceptor cell fate in the eye by participating to the
CC ubiquitination and subsequent proteasomal degradation of Tramtrack
CC (ttk), a general inhibitor of photoreceptor differentiation.
CC Required to block the S phase entry in the peripheral nervous
CC system and central nervous system in a process that does not
CC involve the degradation of ttk. Involved in cone cell development
CC by preventing the transcriptional repression mediated by Su(H) on
CC D1, probably by participating to a E3 complex that contains sno
CC degradation of some component of the Su(H) repressor complex.
CC [1-] SUBUNIT: Component of some E3 complex at least composed of sina,
CC ebi and phyl, required for the degradation of ttk. Probably forms
CC a E3 complex with sno, required for the degradation of some
CC component of the Su(H) repressor complex. Interacts with sno and
CC Su(H) and Smr.
CC [1-] SUBCELLULAR LOCATION: Nuclear.
CC [1-] TISSUE SPECIFICITY: Widely expressed both in embryos and larvae.
CC [1-] DOMAIN: The F-box-like domain is related to the F-box domain, and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes.
CC [1-] MISCELLANEOUS: 'Ebi' means 'shrimp' in Japanese.
CC [1-] SIMILARITY: Belongs to the WD-repeat EBI family.
CC [1-] SIMILARITY: Contains 1 F-box-like domain.
CC [1-] SIMILARITY: Contains 1 Lish domain.
CC [1-] SIMILARITY: Contains 8 WD repeats.
CC [1-] CAUTION: Ref.3 sequence differs from that shown due to a
CC frameshift in position 484.
CC [1-] This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC [1-] EMBL: AF146345; AAD35017.1; --
CC EMBL: AB003589; AAF51501.1; --
CC EMBL: AY061326; AAL28874.1; ALT_FRAME.
CC HSSP: P16649; IERJ.
CC FlyBase: FBgn0023444; ebi.
CC GO: GO:000074; P:regulation of cell cycle; IMP.
CC InterPro: IPR006594; Lish.
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 3.
CC PRINTS: PR00320; GPROTEINRPT.
CC ProDom: PD000018; WD40; 3.
CC SMART: SM00667; Lish; 1.
CC SMART: SM00320; WD40; 8.
CC PROSITE: PS00896; LISH; 1.
CC PROSITE: PS00678; WD_REPEATS_1; 3.
CC PROSITE: PS00882; WD_REPEATS_2; 6.
CC PROSITE: PS00294; WD_REPEATS_REGION; 1.
CC Cell cycle; Nuclear protein; Repeat; Ub1 conjugation pathway; Vision;
CC WD repeat.
FT DOMAIN 4 36 Lish.
FT DOMAIN 41 84 F-box-like.
FT REPEAT 353 392 WD 1.
FT REPEAT 408 447 WD 2.


```
FT REPEAT 449 488 WD 3.
FT REPEAT 491 531 WD 4.
FT REPEAT 532 571 WD 5.
FT REPEAT 574 622 WD 6.
FT REPEAT 625 664 WD 7.
FT REPEAT 666 700 WD 8.
FT DOMAIN 206 323 Ser-rich.
FT MUTAGEN 16 16 L-Q: induces ectopic S phases
within the peripheral and central nervous
system.
FT MUTAGEN 510 510 C-Y: In ebi4; induces defects in R7 cell
development in the eye.
FT MUTAGEN 602 602 S-L: In ebiC3; induces ectopic S phases
within the peripheral and central nervous
system.
SQ SEQUENCE 700 AA; 72387 MW; 28C6D8D07BB79FB7 CRC64;
Query Match 75.5%; Score 2753.9; DB 1; Length 700;
Best Local Similarity 52.8%; Pred. No. 6.6e+47;
Matches 389; Conservative 51; Mismatches 36; Indels 261; Gaps 75;
QY 1 MSISDEVNPLVRYIQESGFHSAPFTGFKSHISQSNINGALVPPAALISIIQKQYV 60
DB 1 MSFSDEVNPLVRYIQESGFHSAPFTGFKSHISQSNINGALVPPAALTIQKGLLYT 60
QY 61 EAEVSINEDGTLFD-GRPIESLSLIDAVMPDV-----VOTRQ-----AYRD---KLA-- 104
DB 61 EVEMSVGEDG---EVARPIEGLSLIDAVMPVKPLPIVKT-EPGKPGAV-DSAP-AGG 114
QY 105 -Q-----Q-----Q-A-----AA--A-----A 112
DB 115 NQNNNAKPEIKIPGTGVAGSGNKIAGTTGTSTPTDQASAEVDSGNAANNAGTVA 174
QY 113 -----A-----A--AA--AQ-----Q-----CSAK--N-G-EN---T-- 130
DB 175 GNNAGGNOASTGSGNSTPTPAGDLAAPGASOKKQNSNEAGSSSGNAGNANATSTD 234
QY 131 -A-----NG-----EE-----N-----G-AH---T- 141
DB 235 AASSTNGNSTSSSVEQPTSGLTAGGTVTSTNPDAASGCASTATGSKAPSGAVTIR 294
QY 142 --A--NN-----H-----T-DMEV--DGD---VEIPPNKAVVLRG 167
DB 295 VQAQGNVQSGSSNAQSSAPSGTISSTSGGAGTAAAL-VPMDIDENIBIPESKARVLRG 353
QY 168 HESEVFICAWNPVS-DLLASGSDSTARINLSE-NSTSGTOLVLRHCIREGQDVPSN 225
DB 354 HESEVFICAWNP-SRDLASGSGDSTARIDWMSDAN-T-NSNQLVLRHCIQGGAEPVSN 410
QY 226 KDVTSLDWNSEGTLLATGSDGPARIW-TKDGNLASTLQGHKGPFPALKWNKK-GNFILS 283
DB 411 KDVTSLDWNCDGSLLATGSDGYARWKT-DGRLASTLQGHKGPFPALKWN-KCNYILS 468
QY 284 AGVDKTTIIWAHTGE-AKQOPFPHPAPALVDVQSNNTFASCTD-MCHVCKLG--QD 339
DB 469 AGVDKTTIIWADTGOCT-QQAFHAPALVDVQNTQAFASCTDQD-IVHCRLGWNE- 525
QY 340 RPIKTFQHTNEVNAIKWPTGNLLASCDDMTLKIMSKQD-NCVHDLQOHNKEITYIK 398
DB 526 -PIKTFQHTNEVNAIKWCPQGGQLLASCDDMTLKIMSMNRDCC-HDLQASHKEITYIK 583
QY 399 WSPGTGTTNNPNANMLASFPSTVRLWDVDRGICHTLTKHOEPPVSVAFSPDGRYLA 458
DB 584 WSPGTGTTNNPNNTLLASASFPSTVRLWDVDRGSCHTLTKKTEPVSVAFPDQKHLA 643
QY 459 SGSFDCVHIWNTQ-----VCLH-YLNGQVLLNLGRSI---CLY-----TLPHLV----- 500
DB 644 SGSFDCVHIWSTQTQOLV--HSY-KG-----TG-GIFEVC-WNSKGT-K---VGASASD 689
QY 501 --VPIPLVALLIYL-K 514
DB 690 GSV--EV-L-D---LRK 699
```

RESULT 10

AAP49011 PRELIMINARY; PRT; 7073 AA.

```
AC AAP49011,
DT 02-MAR-2004 (TReMBLrel. 27, Created)
DT 02-MAR-2004 (TReMBLrel. 27, Last sequence update)
DE Orf1ab polyprotein.
OS SARS coronavirus ZJ01.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus; Human coronavirus (strain SARS).
OX NCBI_TaxID=230471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZJ01b;
RX MEDLINE=22899812; PubMed=14527350;
RA Li L., Wang Z., Lu Y., Bao Q., Chen S., Wu N., Cheng S., Weng J.,
RA Zhang Y., Yan J., Mei L., Wang X., Zhu H., Yu Y., Zhang M., Li M.,
RA Yao J., Lu Q., Yao P., Bo X., Wo J., Wang S., Hu S.;
RT "Severe acute respiratory syndrome-associated coronavirus genotype and
its characterization";
RL Chin. Med. J. 116:1288-1292(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ZJ01b;
RX PubMed=14733771;
RA Wang Z.G., Li L.J., Luo Y., Zhang J.Y., Wang M.Y., Cheng S.Y.,
RA Zhang Y.J., Wang X.M., Lu Y.F., Wu N.P., Mei L.L., Wang Z.X.;
RT "Molecular biological analysis of genotyping and phylogeny of severe
acute respiratory syndrome associated coronavirus.";
RL Chin. Med. J. 117:42-48(2004).
DR EMBL: AY286320; AAP49011.4; -.
KW Polypeptide.
SQ SEQUENCE 7073 AA; 790251 MW; 96595991DACA6EA9 CRC64;
```

Query Match 71.2%; Score 2599.6; DB 2; Length 7073;

Best Local Similarity 8.3%; Pred. No. 1.4e+41; Mismatches 464; Conservative 38; Mismatches 38; Indels 5074; Gaps 381;

```
QY 1 M-----S-----S-----DEV-----N----- 9
DB 297 MGRIRSVYVPASPOECNNMHLSTLAKNCDEVSMTCDFLKATCBHCGTENLVIEGPTT 356
QY 10 ----- 12
DB 357 CGVLPNVAVKPCACQDPEIGPEHSVADYNNHSNIETRLKRGTRCPCGGCVAFVGC 416
QY 13 Y-R-Y-----L-Q--E--S-----G-FSH----- 23
DB 417 YNKRAVVPASADIGSGHTGITGDNVETLNEDLLILSRERNIVGDF-HLNEVAI 475
QY 24 -----SAF-----T----- 27
DB 476 ILASFSASTAFIDTILSKLDYKSFKTIVSCGNYKVTGKPKVGAWNIGQORSVLTPLCG 535
QY 28 F-----G-IKS-----H-----IS-OS-----N-- 38
DB 536 FPSQAGVIRISPARTDAANHSIPDLQAAVITLDGISSEQLRLVDAMVYTSDLTNSV 595
QY 39 -I-----N--GALVP--P-----A-----A-----LIS- 51
DB 596 IIMAVYTGGLVQQTSQLSNLLGTTVEKLRPIPEWIEAKLSAGVEFLKDAWEILKFLITG 655
QY 52 ---I---IQ-----K-----G-----L-----Q--- 58
DB 656 VFDIVKGQIQVASDNIKDCVKCFIDVVKALEMCIQVITAGAKRLSLNLGEVFIQSKG 715
QY 59 -Y-----V-----V-----EA-E--V-S--- 65
DB 716 LYRQCIRGKEQIQLLMLPKAPKEVTFLEGSDHDTVLTSEEVVLKNGEALPTFVDSFTN 775
QY 66 -----I-----N-----ED----- 69
```

Db 776 GAIVGTPVCVGLMLLEIKDEQYCALSPGLLATNNVFRLLKGGAPIKGVTFEGDTWVEVQ 835
QY 70 -----GT-----
Db 836 GYKNVRITFELDERVDRVNLKCSVYTVESGTEVFEPACVBAVAVKTLQVPSDLLTMNG 895
QY 72 -----LFD-G-----R-----P-----I-----
Db 896 IDLEWSVATFYLFDDAGEENFSRMYSFYPPDEBEEDDARCEEBEIDETCSHEYGTED 955
QY 79 -----E-S-----L-----S-----M-----
Db 956 DYQGLPLEFGASAEYVRVEEBEEDWDDTTEQSEIEPEPEPTPEBPVNOFTGYLKLTDN 1015
QY 84 -----I-----D-----A-----V-----
Db 1016 VAIKCVDIVKEAOSANPMWIVNAANTHLKHGGGVAGALNKATNGAMOKESDDYIKLNGPL 1075
QY 89 -----P-----D-----
Db 1076 TVGGSCLLSGHNLAACKLHVGNLNAgedIQLLKAAYENFNSODILLAPLLSAGIFGAK 1135
QY 91 -----V-VOT-R-Q-----QA-Y-----R-----
Db 1136 PLQSLQVCVQTVRTQVYIAVNDKALYEQVYMDYLDNLKPRVEAPKQEEPPNTEDSKTEEK 1195
QY 101 -----D-----D-----KL-A-----Q-----
Db 1196 SVQKPVDRPKIKACIDEVTTTLEETKFLTNKLLLFADINGKLYHDSQNMRLGEDMSFL 1255
QY 106 -----Q-----
Db 1256 EKDPYMGVDVITSGDITCVVIRSKAGGTEMLSRALKKVPVDEYITTPGGCGAGYTL 1315
QY 109 -----A-A-----A-A-----A-A-----
Db 1316 EEAKTALKKCSAFYVLPSPAPNAKEBILGTVSNLREMLAHAEETKLMPICMQDVRAIM 1375
QY 118 AS-Q-----Q-----G-----S-A-K-N-----G-----
Db 1376 ATTQRYKGIKIQEGIVDYGRFFVTSKEPVASITKLNSLNEPLVMPIGYVTHGFNL 1435
QY 128 -----E-----N-----T-A-----
Db 1436 EEAARCMRSLKAPAVSVSPDAVYTYNGYLTSSSTSEHFVETVSLAGSYRDSYSGQ 1495
QY 132 -----NGE-----E-----N-----
Db 1496 RTELGVFEFLKRGDKIVVHTLESVPEPHLDGEVLSDKLKLSLREVKTIKVFTVDNTN 1555
QY 137 -----CA-----
Db 1556 LHTQVDMSTMVYQOQFPTYLDGADVTKIKPHVNHGKTFPVLPSDDLRLSEAFYYHTL 1615
QY 139 -----HT-----
Db 1616 DESFLGRYMSALNHTKKKFPQVGGLTSLKWADNVCYLSVLLAQOLEVKFNAPALQEA 1675
QY 141 -----I-----AN-----N-----H-----
Db 1676 YRARAGDAANFALILAYSNKTGVLGDVRETWTHLLQHANLESASAKRVLNVCCKHGOK 1735
QY 146 T-----D-----MM-----
Db 1736 TTTLTGVEAWMYMGTLISYDNLTGVSIPCVGGRDATQYLVQOBSSSFVMSAPPABYKLOQ 1795
QY 150 -----E-----VDG-----DV-----
Db 1796 GTFLCANEYTGNYQCQGHYTHITAKETLYRIDGAHLTQMSYKGPVTDVFKYKTSYTTIK 1855
QY 156 -----ET-----P-PN-----
Db 156 -----ET-----P-PN-----

Db 1856 PVSUKLDGVITYTEIEPKLDGYKKONAYYTEQPIDLVPTQPLPNASFDNPKLTCSTNPKFA 1915
QY 161 -----K-A-----V-----
Db 1916 DDLNOMTGTGPASRELSVTFPDLNGDVVAIDYRHYSASFKKGAKLLHKPIVHMHNQAT 1975
QY 165 -----LR-----
Db 1976 TKTTFPKNTWCLRLWSTKPDVTSNFEVLAVEDTOGMNLCESQOPTSEEVENPTIQ 2035
QY 167 -----GHE-----S-----E-----
Db 2036 KEVIECDVKTTEVGVNVLKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALG 2095
QY 172 -----VF-----
Db 2096 LKTIATHGIAAINSPWPKILAYVPFLGQAAITTSNCAKRLAORVFNMYVPTLLFQ 2155
QY 174 -----I-----C-----A-W-----
Db 2156 LCTFTKSTNSRIRASLPITIAKNSVKSAKLCLDAGINVYKSPKSLFTIAMWLLLSI 2215
QY 178 -----N-----
Db 2216 CLGSLICVTAAGVLLSNFGAPSYCNGVRELYLNSSNVTTMDPCEGSFPSCICLSGLDSL 2275
QY 179 -----P-----V-S-DL-----LA-S-----G-S-----G-----
Db 2276 DSYPALETIQTVISSYKLDLTILGLAAEWLAYMLFTKFFYLLGLSAIMQVFFGYFASHF 2335
QY 190 -----D-----ST-----A-----
Db 2336 INSNLWMLFIISIVQMAPVSAMVRVYIPFASFYIYHKSIVYHMDGCTSTCMCYKKNRA 2395
QY 194 -----R-----I-----WN-LESEN-ST-SGST-----
Db 2396 TRVECTTIYNGMKRSPVYVANGRGCFKTHNNCL--NCDTFTGTFISDEVARDLSLQ 2453
QY 208 -----Q-L-----V-----L-----BH-----
Db 2454 FKRPINPTDQSSYIVDSVAVKNGALHLYFDKAGQKTYERHPLSHFVNLOLNRANNTKGS 2513
QY 214 -----C-I-----
Db 2514 PINVIVFDGSKCDBSASKSASVYYSQLMCQPIILLDQALVSDVSGDSTEVSKMFDAYVD 2573
QY 216 -----R-----
Db 2574 TFSATFSPWMEKALKALVATAHSELAKGVALDGLSTFVSAARQGVVDVTDVTDKDVIECLK 2633
QY 217 -----E-G-----G-----Q-----
Db 2634 LSHSDLEVTGDSCNNFMITYNKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVK 2693
QY 221 -----
Db 2694 DYNLSLEQLKQIRSAAKKNIPFRITCATTRQVNVNITTKISLKGKIVSTCFKLMLKA 2753
QY 222 -----V-P-----S-----N-----K-----D-VT-----S-----LD-232
Db 2754 TLLCVLAALVCYIVMPVHTLSIHDGYTNBIIGYKATODGVTRDIISTDDCFANKHAGFDA 2813
QY 233 W-----N-S-----E-----GT-----
Db 2814 WFSQRGSYKDKSCPVAAIITREIGFVPGVLPGLVLRINGDFLHFLPRVPSAVGNIC 2873
QY 239 -----I-----LA-----T-----G-----SY-----
Db 2874 YTPSKLIEYDFATSACVLAECTIFPKDAMGKPVPCYDYNLLEGISYSSELPRDTRYVL 2933
QY 246 -----DG-----F-A-----R-----I-----W-----
Db 2934 MDGSIIFPNTYLEGSRVVTTFDAFYCRHGTCERSEVGICLSTSGRWVNLNNEHYRALSG 2993

QY 253 -----TK-----D----- 255
Db 2994 VFCGVDAWNLIANIPTPLQVQALDVSASVWAGGIIAIIIVTCAAYYPKFRFRVGEYNH 3053
QY 256 -----G-----N-----LA-----S----- 260
Db 3054 VVAANALLFLMSFTILCLVPAYSFLGVSVFYLTYFTFTNDVSLAHLQWPAFSPIV 3113
QY 261 -----T-----L-----G----- 263
Db 3114 PFWITAIYVFCISLKHCHWFFNYLKRVMFNGVTFSTPEEAALCTFLNKEMYLKRSE 3173
QY 264 -----Q-----H-----K-----G-----P----- 268
Db 3174 TLLPLTOYNYRLALYNKYKYPFGALDTSYREAAACHLAKALNDFNSGADVLYQFPQTS 3233
QY 269 I-----F-----AL-----K-----W----- 274
Db 3234 ITSAVLQSGFRKMAFPKGVGCMVQVTCGTTTLNGLMLDDTVYCPRHVICTAEDMLNPN 3293
QY 275 -----N-----K----- 276
Db 3294 YEDLLIRKNSHSLVQAGNVQLRVIGHSMQNCILLRLKVDTSNPKTKYKPVRIQPGQTS 3353
QY 277 -----KG-----N-----F-----I-----L-----S-----A 284
Db 3354 VLACYNPSGVYQCAMRPNHTIKGSLNGSCSVGFNDYDCVSCYMHMELPTGVHA 3413
QY 285 G-----VDK-----TTI-----I-----W-----D----- 294
Db 3414 GTDLEKFGYGFVDRTQAAGTDTITLNVLAWYAAVINGDRWFLNRTTLNDFNLV 3473
QY 295 A-----H-----TG-----E----- 299
Db 3474 AMKYNEPLTQDHVIDILGPSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTP 3533
QY 300 -----AK-----Q-----O-----F-----PF----- 306
Db 3534 FDVVRQSGVTFOGKEKIVKGTTHWMLLTFLTSLILVQSTOWSLPFFVYENAFIPFTL 3593
QY 307 -----H-----S-----A-----PA----- 311
Db 3594 GIMAIACAMLLVKHGHAFCLFLPLSLATVAYFNVMYVPASVMRMIMTWLEADTSLG 3653
QY 312 --L-D-----V-D-----W-----O----- 317
Db 3654 YRLKDCMYASALVLLIIMTARTVYDDAARRVWTLNMVITLVKYVYGNALDOAISMWAL 3713
QY 318 -----SN-----NT----- 321
Db 3714 VISVTSYSGVTTIMFLARAIUVFCVVEYYPPLLFTIGNTLQCLIMVYCFGLYCCCYFGL 3773
QY 322 FASC-----ST-----D----- 328
Db 3774 F--CLLNRYFRLTLGVVDVLSVSTQEPYMYNSQGLPPKSSIDAFKLNILGIGGKPCIK 3831
QY 329 -----M-----CI-----H----- 332
Db 3832 VATVQSKMSDVKTSVLLSVLQQLRVESSSKLWQCQLHNDILLAKDTTEAFERQVSL 3891
QY 333 -----V-----CK-----L-----G----- 337
Db 3892 LSVLLSMQGANVINRICEMLNDRATLOIASEFSLPSYAAYATAQEAQEAVANGDSE 3951
QY 338 -----Q----- 338
Db 3952 VVLKKLKSLNVAKSEFDDAAMQKLEKMAQAMTQMYKQARSEDKRAKVTSMQTMFLP 4011
QY 339 -----D-----R-----P-----I-----K-----TP----- 345
Db 4012 TMLRKLDNALNIIINNARDGCVPLNIIPLTTAAKLMVVVVPDYGTYKNTCDGNTFTYASA 4071

QY 346 -----Q-----GH-----T-----NE-----V----- 352
Db 4072 LWEIFQVVDADSKHVQLSEINMDSNPLAWPLIVTALRANSVAVKLQNNELSPVALROMSC 4131
QY 353 -----NA-----I-----KW-----D-----P----- 359
Db 4132 AAGTTQTACTDDNALAYYNNNSKGRFVLLALSDHQLDKWARFPKSDGTGTIYTELBPPCR 4191
QY 360 -----T-----G-----NL-----LA-----S-----CS----- 368
Db 4192 FVTDTPKGPKVLYPIKGLNNLRGMLGSLAATVRLQAGNATEVPAHSTVLSFCAPAV 4251
QY 369 D-----D-----M-----T----- 372
Db 4252 DPAKAYKDYLASGGQPIITNCVKMLCTHTGTGOAITVTPEANMDQESFGGASCCLYCRCHI 4311
QY 373 -----LK-----I-----W-----S-----M----- 378
Db 4312 DHPNPKGCDLKGKYYQIPTTCANDPVGFTLRNTVCTCGMMKGYGSCDQLREPLMQSA 4371
QY 379 -----K----- 380
Db 4372 DASTFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNEKAVGPAFLKTNCCRFQEKDEE 4431
QY 381 -----D-----N-----C-----VHDL-----Q----- 388
Db 4432 GNLLDSYFVVKRHTMSNYQHEETIYNLVKDCPAVAVHDFKFRVDCGDMVPHISRQRLTKY 4491
QY 389 -----OH-----N-----KEI----- 394
Db 4492 TMADLVYALRHFDEGNCDTLKEILVTYNCCDDYFNKKDWDYFVENPDILRVYANLGERV 4551
QY 395 -----Y----- 395
Db 4552 ROSLLKTVOFCAMRDAGIVGVLTLDNQLNGWYDFGFEQVAPCGGVPIVDYSYSLLM 4611
QY 396 -----T-----IKW----- 399
Db 4612 PILTLTRALAASHMDADLAKPLIKDLKLYDFTBEERLCFLDRYFKYWDQTYHPNCINCL 4671
QY 400 -----S-----PT-----GP----- 404
Db 4672 DDRCILHCANFNVLSTVPFPTSPGFLVRKI FVDGVFPFVSTGYHFRELGVHVNQDNLH 4731
QY 405 -----G-----TNN-----P----- 409
Db 4732 SSRLSFKELLVYAADPAMHAASGNLLDKRTCTCFVSAALTNVAVFOTVPGNFKNDFYDF 4791
QY 410 -----NA-----NL-----M-----L----- 415
Db 4792 AVSKGFFKEGSSVELKHPFFAQDGNAAISDYDYRYNLTMTCDIRQLLFVVEVVDKYFDC 4851
QY 416 -----A-----SASF-----DS----- 422
Db 4852 YDGGCINANQVNNLDKSGAPFNKWKARLYDSMSYEDQDALPAYTKRNVIPITLML 4911
QY 423 -----TV-----R-----L-----W----- 427
Db 4912 NLKYAISAKNRARTVAGVISCSTMTNRQHQKLLKSIAATRGATVVIGTSKFGYGGWHNML 4971
QY 428 -----DV-----DR----- 431
Db 4972 KTVYSDVETPHLMGWDPYKCDRAMPNMLRIMASLVLARKHNTCCNLSHRFPYLANECAQV 5031
QY 432 -----G-----IC-----I----- 435
Db 5032 LSEMVMCGSLYVPGGTSGDATTAYANSVENICOAVTANVALLSTDGNKIADKYVRN 5091
QY 436 -----H-----TLT-----KH----- 441
Db 5092 LOHRLYECLYRNRDVEHFYAYLRKHFSMMLSDDAVVCYNSNYAAQGLVASIKNF 5151
QY 442 -----Q-----E-----P-----VY-----S-V-----A----- 449

Db 5152 KAVLYQNNVFMSEAKWTETDLTKGHEFCQSQHTMLVKQDDYVLPYPDPRIILGAGC 5211
QY 450 -----F-S-----P-----D-----G-----454
Db 5212 FVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNOEYADVPHLYLQYIRKLHDELTHGLMDM 5271
QY 455 -----RY-----L-----A-----S-----GS-----P-----462
Db 5272 YSVMLTNDTSRIWEPEFYBAMTPHTVLOAGACVLCNSQTSLRGACIRRPFLCKKC 5331
QY 463 -D-----K-----C-----Y-----H-----467
Db 5332 YDHVISTSHKLVSNVPYCNAPGCDVDTQVLVIGMSYYCKHPPIPFPLCANGQVF 5391
QY 468 -----RY-----I-----W-----NTQVC-----474
Db 5392 GLYKNTCVGSDNVDFNAIATCDWTNAGDYILANT--CTERLKLFAAETLKATEETFKLS 5449
QY 475 -----LH-----476
Db 5450 YGIATVREVLSDELHLHLSVEGKPRPLNRNYVFTGYRVTKNSKVQIGEVTFEKGVDGDA 5509
QY 477 -----Y-LN-G-----Q-V-----L-----LN-----485
Db 5510 VVYRGTTTKLVNDGYFVLTSHVTMPLSAPTLVPQEHYVRIITGLYPTLNLSDFFSSNVAN 5569
QY 486 -----L-----G-----R-----488
Db 5570 YQKVGQKYSTQGGPGTQKSHFAIGLALYPSARIIVYTACSHAADVADALCEKALKYLPID 5629
QY 489 --S-I-----C-----L-Y-----T-----LP-----496
Db 5630 KCSRIIPARAVECFDKFKNSTLEQVFTVNALPETTADIIVFDEISMATNYDLSVN 5689
QY 497 -----H-----497
Db 5690 ARLRAKHVYVIGDPAQLPAPRTLLTKGTLPEYFNSVCLMKTIGPDMFLGTCRCCPAEI 5749
QY 498 -----H-----L-----V-----501
Db 5750 VDTVSALVDYNDKLKAHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVVRFLTRNPAWRK 5809
QY 502 --I-PL-----VA-----L-----I-----EL-LVL-----513
Db 5810 AVFISPINSONAVASKILGPTQTVDSQSGSEYDVIPTOTTETAHSCNVNRNVAITRA 5869
QY 514 K 514
Db 5870 K 5870

RESULT 11
Q8NJX1
ID Q8NJX1 PRELIMINARY; PRT: 20925 AA.
AC Q8NJX1
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Nonribosomal peptide synthetase.
GN Name=txl;
OS Trichoderma virens.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=29875;
RN [1]
RP MEDLINE=22037864; PubMed=11909873;
RX Wiest A., Grzegorski D., Xu B.-W., Goulard C., Rebuffat S.,
RA Ebole D.J., Bodo B., Kenerley C.;
RT "Identification of peptaibols from Trichoderma virens and cloning of a
RT peptaibol synthetase";
RL J. Biol. Chem. 277:20862-20868(2002).

RN [2]
RP SEQUENCE FROM N.A.
RA Wiest A.S., Grzegorski D., Xu B.-W., Zhou W., Ebole D.J.,
RA Kenerley C.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL: AF469045; AAM78457.1; --
DR HSSP: O30409; 1DNY.
DR GO: GO:0048037; F:cofactor binding; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR010071; AA adenyl_dom.
DR InterPro: IPR009081; ACP_like.
DR InterPro: IPR001227; AC_transferase.
DR InterPro: IPR000873; AMP_bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR000794; Ketoacyl_synth.
DR InterPro: IPR006162; Ppantne_S.
DR InterPro: IPR006163; Pp_bind.
DR InterPro: IPR010080; Thioester_redct.
DR Pfam: PF00698; Acyl_transf_1; 1.
DR Pfam: PF00501; AMP-binding; 18.
DR Pfam: PF00668; Condensation; 27.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF00550; PP-binding; 19.
DR PRINTS; PRO0154; AMPBINDING.
DR TIGRFAMs; TIGR01733; AA-adenyl-dom; 17.
DR TIGRFAMs; TIGR01746; Thioester_redct; 1.
DR PROSITE; P800075; ACP DOMAIN; 19.
DR PROSITE; P800455; AMP_BINDING; 18.
DR PROSITE; P800012; PHOSPHOPANTHETHEINE; UNKNOWN 11.
SQ SEQUENCE 20925 AA; 2320297 MW; C6603A269D56FE11 CRC64;

Query Match 71.2%; Score 2598.3; DB 2; Length 20925;
Best Local Similarity 8.2%; Pred. No. 1.2e-40;
Matches 465; Conservative 41; Mismatches 6; Indels 5177; Gaps 385;

QY 1 M-----S-SS-----D-----E-----V-----8
Db 8843 MISYEQYGLQNTSKISSHAMEACDFGLIIQIEQLAYRGDDALLVAEASHITNETSAL 8902
QY 9 -N-F-----LVY-----R-----Y-----LOE-----18
Db 8903 QNVFSPVLIQCHLYDNTNLMLIYDSQALSSRMEALSHHPDNVQVELASKDNLLLSI 8962
QY 19 S-G--F-----S-----H-----S-A-----FT-----27
Db 8963 SLAGRWDFFQALKLNSDSPDAVERCFHEMLDEIALVRGDSLAIISGWDKSFYREMAETTN 9022
QY 28 -----F-----G-----I-KSH-----I--34
Db 9023 RIAHLVTDYGVKVGDIHVCFEKSAMFTIATLAINKAGAAWSTLDPSPHPTERYQKIINQ 9082
QY 35 --SQ-----SN-----46
Db 9083 TGSQIALSSPVNSRCASLLPNVIELTSELDSKLAQNVSWTSRPAVNTPRDAAVLFT 9142
QY 39 -----I-----N-----GA-42
Db 9143 SGSTGVKGVWIEHGSLSCTSIQSLSKALDFNEFRVLQFSSYFDFIILFEIGSTFLTGC 9202
QY 43 -----L-----VP-----P-----46
Db 9203 LFVPSWDEQNMELVEYIRKHQLTFFMLTPTLTARTIRPEDVPSVDMVLVAGEAPTRDILDI 9262
QY 47 -----A-----A-L-----I--S-----51
Db 9263 WFGKRLANGWGTSCSVIACLHQWTSVDESPKVIQRPICGGSCWIVDPEDATCWAPLGTV 9322
QY 52 --I-IQ-----K-----GL-Q-----Y-----V-EA--E 63

```
Db 9323 GEIVIOQRNLLREYLSDPVKATATITGLPQWPKRDSIHWDRFYLTGDLGFINEAGNLE 9382
QY 64 -----VSI-----N-E-----D-G-TL----- 72
Db 9383 YCTRKTQVKIRQORLEGEIEHHIQANLESVRQAVDVVTKSDAGSTLVAFVSFSDATQ 9442
QY 73 -----F---DG-----R---P---I----- 78
Db 9443 TVTDTRVIANGPSTLDGDFQATISRLIGILSTLMPRYMVPASAFIPCAVKGFEQHTGGRY 9502
QY 79 -----E---SLS-----L---ID----- 85
Db 9503 KKLRELAASQEDMAAYSLANBEKVAPOTAMESRIQKIWAQVLNISIDSIGRDDSFLQI 9562
QY 86 -----AV-----M---P---D--- 90
Db 9563 GGDVSUVRILSVARDANIRLTVDIFDDPRLIAVAKATEDWRDGLMTQIEPFFVLDDG 9622
QY 91 -----VV---QT-R-----Q-----A---YR-----D 101
Db 9623 SVKNIVLSQNIREQYSLPADKELEDAYPCTKLQEGLMALAVKQPGSVIAKFLYRISPNVD 9682
QY 102 -----K-----L-----AQ-----Q-----QA 108
Db 9683 VSHLRASWEETVRLPLNLRTRIFTVNMSVQAVFKDDVSNQSTSGTSLRSYMCGSTENFQM 9742
QY 109 A-----A---A-----A-----AA----- 114
Db 9743 AYGSPLARVALIEDGGHTFYILSMHHAUVFGWGMVFMVAFHISIFHNRTHTVVEPVRFV 9802
QY 115 -----A---AAA---SQ-----Q----- 121
Db 9803 KYTLAIDSDAAADYMSQEGARQSIFPPTLSSVHEKKENSTLTLEKVIDLPSTKSSIT 9862
QY 122 -----G---S-----A-K----- 125
Db 9863 TATMLRAAWAAILARYCSDSDITFGATISGRQAPVFLADMTGPAITVPRVYVIRNQL 9922
QY 126 -----N-----G----- 129
Db 9923 ISDYLOAIQRANDMTIPFQFGLQNTSKLSDDAKADCFSSLLVIOPIQSLSVYDENADA 9982
QY 130 -----T-AN-----G---E-----EN----- 136
Db 9983 IFVQAEVEKEIGTVQNYFSYPLVIOGHVENFINLVLYDSNVLSNQITALTSHQFESV 10042
QY 137 -----G---A-----H-----T-----I--- 141
Db 10043 MKQLASQSDMKLGSVMSASDWDLEHSMRQNRDIPDIVDSCIHELIOYQAVTQPDAPAIVS 10102
QY 142 -----A-----A-----N-----N----- 144
Db 10103 WDRDFTYKQLEASRNLALHLVNKNYKPKDDLIPVFFERKAMVFAITAINKAGAAWVPL 10162
QY 145 -----H----- 146
Db 10163 DPSPHVLRLRQLISQGTTLALSSANAVLCSTLVRKVVEVNAELDNKLLATESAHGPV 10222
QY 147 D-----M----- 148
Db 10223 VDVSSRNAAYVLFTSGTGIPKGLIMEHGSVCTSQVAIAKRLGLNSKVRILQFAAFVFDL 10282
QY 149 ----- 148
Db 10283 SIGEIVGLISGACICVPSEHIRKNSIANFINRQGITWTYLTPTSFVRTIKASEVNPVKLL 10342
QY 149 -----M-----E-----V 151
Db 10343 LLAGAVPRDIFATWFGKRLRLINGWPAETCCFSTLHEMAICGRESFYCCPRPVGVCWIV 10402
QY 152 D-----G-----DV---E---I---P---P-----NK 161
Db 10403 DPENPHRLAPTGAIGEVIIOGPTILREYLSVDVORTEAAVIKSLPDWAPFREQSSGRFNK 10462
```

```
QY 162 -----AV-V-L----- 165
Db 10463 SSDLGVYNPDGTIEFSSRKDTQVKIRGLRVELGEVEHVAQVALDGVHQIAVDVFKGDNGT 10522
QY 166 -----RG-----HE---S-----E-----V-----FI-C-- 175
Db 10523 QPRGYFSDESQIHEAHPSGFPQAIDEKQLARLTVAIGKLNIALPRYMIPTLFPCKY 10582
QY 176 -----A----- 177
Db 10583 MFSITSTKLDRELNRRTLLTQSELAMFSYRCKNKTPTPTMESLQRIWSEILHPLDS 10642
QY 178 -----NP-----V-----S 181
Db 10643 IGRDSDFLGLGGDSITAIHLGSISSRDQGISITAKOVFDDPRLWDIASKARELDEAQQQS 10702
QY 182 D-----L-----LASGS-----G----- 189
Db 10703 PDISPFSFLIGETRELAAGSVKQKLKLTQSOTIEDAYPCTSKIPIGSIYAKYVYRPLHQV 10762
QY 190 D-----S---T-A-----RI-----W-NL-S---E-N-----S 202
Db 10763 DIGRFKSSWEHTVALHPILRTRIIMIKOTCVQLVVKDSVTWENLDSGDLDEAVNATHSYS 10822
QY 203 -T-----SG-----S- 206
Db 10823 MTYGSPLSRYSIHESKSGDKYFLWTAHHSIHGWSVPVIENTLYQAYKGLGKPKAYS 10882
QY 207 -T-----OL----- 209
Db 10883 FIKYTWMEHQKAAGDYWRQLOQNAKASFPKGTGLAASIKWTGRKTRVMSTSLKFSASSNE 10942
QY 210 -----V---L-----R----- 212
Db 10943 IATKASVIRAAWAVILARYCESDDVCFGATISGRQASVFCLEMGAPAVATVPVRVOLDN 11002
QY 213 -----H-----C-----I-R----- 216
Db 11003 DQESKFLQNIQSOAHMVPYEQYGLQSTAKLGNARDARDCTSLVLIQVIOQRFESVQS 11062
QY 217 -----E-GG-----Q---D-----V-----P-- 223
Db 11123 IVQQLAAEDGGKLGDISIASSFDLELAQRSNGDGPQIIDDCHIHLIERQAKORPNRPAI 11182
QY 224 -----S-----NK----- 226
Db 11183 DAWDGKFTYSELDRTRANLRLAHLVHDYAVKVGDIHVHCFEKSQWYFVAILAVNKAGAWA 11242
QY 227 -D-----V-----T-----S-LD-W---NSE- 236
Db 11243 PFDPAHPQRLKAVASQTCAGKALASTANTRLCEQVDCVWEVSSTLDKNLWTTYNIEK 11302
QY 237 G-----T-----L-----A-----T----- 242
Db 11303 GPDINVTMPDAAAYILFTSGTGVPKGIVMQHGAICTNQAAALSGWLGFHTVRMLQFSSFV 11362
QY 243 -----G-----S-----Y----- 245
Db 11363 FDSVGEIVQALMNGACVCPSEHRLNLSDFVRDPNVTWAYLTPTSFRTTLKPKDPPSL 11422
QY 246 -----D---G---F-----A-----R----- 250
Db 11423 KULLAGEPTQDVLDTWFGLEPNTRFINAWGPAETCVYNTLYEWQSNSTESPLKLGAVGA 11482
QY 251 -IW----- 252
Db 11483 YIWWVDVENPORLAPTGCIGLIIVQGPPLLKEYLADPEKTAATAATVTEPEWAPRQSTTW 11542
```

QY 253 ---T---KD----- 255
Db 11543 NRFRDGLGVDHGMHFAASKDQTKIRGLRVELGEVEHRIQGLDGRQVAVDVFK 11602
QY 256 ---G-NL---AS--T---LG- 263
Db 11603 TERGANLVAYLCTSDTKTPDQNTDFGKDVFPASIDFELQINLGKMLTELNSSIPTWIP 11662
QY 264 ---Q----- 264
Db 11663 TMFIPCAYPEFITSSKLDRLVRLTAELSQQOLEAYSLLVNEKQAPETEMEIRLOKMA 11722
QY 265 ---H---K-G- 267
Db 11723 EVLDLPEASIGRDNFMRIGDGSIAAIRLVSMARDAGISLTVNDIFDDARLISVATKAID 11782
QY 268 ---PI---FAL- 272
Db 11783 NDESHLWAPIEPFSLTSGIEDFALPSELSDGQFIEDAYPCSKLOEGLMALAKQPGSY 11842
QY 273 ---K-W---N- 275
Db 11843 IAKYIYKLSHVVDKFKAAWESTVEIASALRTRIIQTGGCSIQVWINGDISWDEADGDL 11902
QY 276 ---KKG---N--F---I--LS---AGV 286
Db 11903 QSLAQSOQLWGYGTRLCRYTLIRRDNDTYFLWMMHVAVIDGLSTQNIILGTLFNIYSGV 11962
QY 287 ---DK- 288
Db 11963 DVLPPLPNRFIOYMLQLEEDTANYWKNLHNAQRTIFPPNDVARDKPAATQMLQSSIE 12022
QY 289 ---T-TI---I--W- 293
Db 12023 LPHGLDNSTVTIATVIRAAWAIVLARYCDSDDVTFGTTISGROAPIEIMGVGTIATV 12082
QY 294 ---DA- 295
Db 12083 PVRVLNRQQLISDFLEGVORQAVEMIAFQYGLQNTAKLGGDARDACDFSSLLVQPIQ 12142
QY 296 H---T---GE-A-K-Q- 304
Db 12143 HLAGTKDDSLVDASIGEDNAVEAIQNYFSPYPLVIAHLYDDHISFILIYDSSIMLEPQ 12202
QY 305 ---P- 306
Db 12203 MVALSVHLQVMMQLTNTTPATLETVSVSSSYDVERALAINAEIPEVIDTCHEMFENRV 12262
QY 307 H-S--APA- 315
Db 12263 HLNPLAPAAVADGNYYTAEBSAANKLAHLYKSYSVKLNDLVHVCDFKSAWYIVSIIA 12322
QY 316 ---W---Q---S-N---N- 320
Db 12323 INKAGATWVLPDPHPEQRLSIVNQTKTALASPGNITLCSALVDNVVVESSALDSML 12382
QY 321 ---T---FAS- 325
Db 12383 PAQGLSPKSVSSRTAAVILFTSGTGTPKGVVIOHRSRLCTNMNAIAKRVRYHLDVRI 12442
QY 326 ---ST--D--M---C- 330
Db 12443 LQFSAYVDFDSILEIINSLQACVCPVSEHIRMNGIVDFIRDMNINWLTPSLRTIN 12502
QY 331 ---IH- 332
Db 12503 PIDVPNVELVAGGEAIPRDVFETWGRVRFINGWGTETTVVGSIHESFESVDESPSTIG 12562
QY 333 ---V--C---K--LG- 341
Db 12563 HPVGGFCWIVDPNPNQLAPTGTGLGEIVIOGPTLLHLYLNDPKTOBAILYDLPEWAPRP 12622
QY 342 ---IKT---F- 346

Db 12623 DENNWGRPYKTDGLGFYNANGKIEFSRKDTQVKIRGLRVELGEIEYQVQASVEEIRQIA 12682
QY 347 ---G---H---TN--E---V- 352
Db 12683 VDVIKTDGNSNLVAYLCFNDIEIROLHNADVNGPSPDLTNLQETLAGAIGKLSVTLPRYM 12742
QY 353 ---N---A---I 355
Db 12743 IPTFYPCSYMPSITSGKLDRLKELKQTAALSOSELNKFSLHGVSKRAPETPMELQONI 12802
QY 356 KW- 359
Db 12803 -WSKLLSIPSESIGRDDSFLGLGDSIMAIHLVTACREAGVSLTVKEIFDDPRLSAVASH 12861
QY 360 ---T- 360
Db 12862 ARSMDAIDQDLVLPFLLSRLREMLVSDTRFQTSLLPSQVIEDAYPCSKLOEGLMAL 12921
QY 361 ---G---N---LL- 364
Db 12922 SVKQSGYVAQYVYKLSGVDLVRFKASWERTIQLCANLRTRIVMLDGTVCQLLVDPAPAE 12981
QY 365 ---A-S--C---S---D---D---M- 371
Db 12982 WDNAPSTDQTLNATRESEMYGSRNLRYALVRDPFEGNHFIVWSSHAVHDGWTLRIMN 13041
QY 372 TL- 373
Db 13042 TLYSLYDQGLPNLLPYSAPIRYTVNIDKEASRFWTEQMKNAKRATYPPMPRIESHTGV 13101
QY 374 ---K-I--W- 378
Db 13102 SRMMNMTISFPSPSVKTVNTKATILRTGTAAILARYCDTDDITFGTTVSGRQAPVGLTEM 13161
QY 379 K---Q---D---N---C- 383
Db 13162 PGPVATVPIRVLEASQIVAOFLSNIQSOATDMIAFEQGLQNI1K1STDADACEFSS 13221
QY 384 -V---H-DL- 388
Db 13222 LLVIQPRSHDLTIKGSESSELVAVSADARSQEQLIYFTYPLVIQGHVFEDSIELLT 13281
QY 389 ---Q---H---N---KE- 393
Db 13282 YDSTILSEVQKALSHQFNVVANQLVNESNDPLSSITISGEWDLQAKENWVENPELDT 13341
QY 394 ---I---YT- 398
Db 13342 CIHSLIEKQARIRPDAPICANDGENYQNLNSAANQLAHILKIGIKADDDLHVVCFEKS 13401
QY 399 -W--S---PT---G---POTN- 408
Db 13402 VMFVSIIAINKVGAADVPLDPSHPQRLRVQVVGOTLAKFALSSP-TNAALCNKLVHVI 13460
QY 409 ---P---N-A---N---L-M- 416
Db 13461 EVSPSLIDLSKFCDFGNFPAINVPSSNAYVLFTSGSTGTPKGLVMQHGAVCTSQTAIA 13520
QY 417 ---S---AS--FD- 427
Db 13521 KRLSLTPDVRILOFAAVVFDLSIGEIVAPLIHGACVCPVSEETRMNGLKEFIRDAWINWA 13580
QY 428 ---DV- 429
Db 13581 YLTPSFVTRLRPEDVPSLQALLLAGEAVGRDILDWTWFGKVRKLINGWPAETCVFSTLHEW 13640
QY 430 ---D---R--G- 434
Db 13641 SSIDESPLTIGRPVGYCHIVEAEDSNKLTPIGCLGEVVLQVFTLLREYLADPQRSKETI 13700
QY 435 I---H- 440

Db 13701 ITPLPWPAPKQVSDAHLWSRFYKSGDLCPYNPNGTLEFFYSRKDTQVKIRGLRVELGEVEH 13760
QY 441 H-----Q-----E-----P-----VY-----446
Db 13761 HIRELLEGVRQAVDVLITSETGTQLVSVYICFNDDSQSPSPKASDIYIPLDADIQARIT 13820
QY 447 -----SVA-----F-----450
Db 13821 SMVGSLSVTLPRYMIPTLFIPCKYMPFITSTKLDRTKLSITASLGRDELIHVSLSNSKK 13880
QY 451 -----S-----P-D-GR-----YLAS-----G-S-----461
Db 13881 RVPETEMETRLQAIWSEILNPLDSTIGRDDSFLQIGDSDITAIYLSKAREAGISLIUKD 13940
QY 462 -FD-----K-C-V-----HI-----468
Db 13941 IFDDSELLAVASKAVLSTEFQBPQPIVFPFSLNKNKTRALVLAGEVRKFCGLNDHDIETD 14000
QY 469 -----WN-----470
Db 14001 AYPCTSLQELMALTMKQPGSVYAKYVYKLSFVDMERFQAANRTMELCGNMRTRIVLL 14060
QY 471 -T-----Q-----V-----C-----L-----475
Db 14061 NGTPIQLLKEDNQWOLENDTLASITNSSRDLKMGYGAPLCWYGVLENNNNKYFWWSAH 14120
QY 476 H-----YL-----N-----G-----480
Db 14121 HSIYDGNWVRILLTYLSTDTVPTLPYSGFIKYNMELDSMALADFWRELQSGSKRA 14180
QY 481 -----Q-----VL-----483
Db 14181 VFPARTSPSSSSSTQIPKSIISLEQAKQSIIITKASILRAAWAIVLARYCDTDDVSFGTT 14240
QY 484 -----LN-----L-----G-----487
Db 14241 VSGRHAPVAGLETMPGPMIATVPVRVHLNRSATKSQFLTEIQOAYEMVYVPEQFGLONIS 14300
QY 488 -----RSIC-----L-Y-----T-----LP-----496
Db 14301 KLSRDARTCDFSSLLVIQPPATTISEKADTNILYIGDEBQSLLTDAMHNYFNPLVII 14360
QY 497 -----H-H-----LV-----500
Db 14361 LNTFEDHILQRFENSEVLTEARVSALSNDHVDVKKLAETDETLESVNLGVDWDIQHA 14420
QY 501 -----V-----501
Db 14421 LASTRLKPSTESTHWLIQERIKTPQNDTAIVSWDGLTYKELGVLASRLAWKLOGLGVG 14480
QY 502 -----IPL-----VAL-----L-----V-----L-----K 514
Db 14481 PESLIPLCPKSTWAVAMVAIEMAGGAFVPLDPNAPVARLGRITIEDTK 14529

RESULT 12

Q96RW7 PRELIMINARY; PRT; 5636 AA.
AC Q96RW7
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hemictentin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Trent J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156100; AAK68690.1; --
DR HSSP; P07996; 1LSL.

DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005021; F:vascular endothelial growth factor receptor. .; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000152; Asx hydroxyl_S.
DR InterPro; IPR000875; Cecropin.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR006605; G2F.
DR InterPro; IPR009030; Grow fac_recept.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig c2.
DR InterPro; IPR000169; Pept_cys_acsite.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR009134; VEGFR.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF07645; EGF_CA; 8.
DR Pfam; PF07474; G2F; 1.
DR Pfam; PF00047; IG; 44.
DR Pfam; PF00090; TSP 1; 6.
DR PRINTS; PR01832; VEGFRECEPTOR.
DR SMART; SM00179; EGF_CA; 7.
DR SMART; SM00408; IGC2; 43.
DR SMART; SM00209; TSP1; 6.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX HYDROXYL; 5.
DR PROSITE; PS00268; CECROPIN; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS00026; EGF_3; 5.
DR PROSITE; PS01187; EGF_CA; 8.
DR PROSITE; PS00835; IG LIKE; 44.
DR PROSITE; PS00639; THIOLE PROTEASE_HIS; UNKNOWN_1.
DR PROSITE; PS00092; TSP1; 6.
DR EGF-like domain.
KW SEQUENCE 5636 AA; 613673 MW; F000B319CED7B52C CRC64;

Query Match 71.2%; Score 2596.5; DB 2; Length 5636;
Best Local Similarity 9.7%; Pred. No. 1e-41;
Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;
QY 1 M-S-----I-----SS-----D-----EV-----8
Db 1 MISWEVHTVFLFALLYSLSLAQDASPQSEIRAEFPPEGASTLAFVDFVTGSMYDDLQVVI 60
QY 9 -----NF-LV-----12
Db 61 EGASKILETSLKRPKRPLNFALVPFHDPEIGFVTTITDTPKPFQVELRELYVQGGDCPE 120
QY 13 -----Y-----R-----Y-----LQ-----17
Db 121 MSIGAIIKALEISLPGSFYVFTDARSKDYRLTHEVLQIQKQSQVVFVLTGDCDDRTH 180
QY 18 -----E-----SG-----F-----S-----H-----S-----A-----F 26
Db 181 IGYKVEIEIATSSSQVFHLDKKQVNEVLKWVEEAQASKVHLLSTDHLEQAVNTWRIF 240
QY 27 -----T-----I-----PG-----I-----K 31
Db 241 DPSLKEVTVLSGSPMIEIRNPLGLIKKGFLHLLNIHNSAKVNVYKEPAGMWTYK 300
QY 32 -----S-----H-----I-----S-----Q-----36
Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPFLDFKKTVSRPVQGIPTVYLLNTSGISTPARID 360
QY 37 -----S-----S-----NI-----N-----G-----41
Db 361 LLELLSISGSSLSKTIPTVKYYPHRYKPGIWNISDFVPPNEAFPLKVTGYDKDYLFORVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
Db 421 VSFSSIVDPADPKVTMPKTPGYYLQPGQIPCSVDSLLPFTLSFVRNGVTLGVDQYLKESA 480

Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIPIINKGDLWGPGLSPKEVKIKVNTLTLECE 2698
QY 300 A-----K-----Q-----Q----- 303
Db 2699 AYAIPSASLSWYKGGQPLKSDHVNIAANGHTIQLKEAQISDGTGYTCVASNIAGEDEL 2758
QY 304 F-----P-----F-----H----- 307
Db 2759 FVNIQVPPSFQKLWEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPPTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LITSKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPIIEGANS 2878
QY 316 -----WO-----SN----- 319
Db 2879 LPEEVTVLNKSALIECLSGSPAPRSMQKQDQPLEDDHKKFLSLGRILQILNTQITD 2938
QY 320 -----NT-----FAS--C-----S 326
Db 2939 IGRYVCVAENTAGSAKKYFNLNVLVPPSVIGPKSENLTVVNNFISLTCVSGFPFPPDL 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKNQOPIKLTNTLIVPGGRTLQIIRAKVSDGGEYTCIAXNAGSCKKFSLTIVVPPS 3058
QY 332 --H-----V--C-----K-----L--GO----- 338
Db 3059 IKDHDSESLSVNVREGTSVLECESNVPPVITWYKGRMITESTHVEILLADGQMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KKAESVDTQYVCRAINVAGRDDKNFNLVNVPPSVIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 --T-----P-----Q-----G-----H-----T----- 349
Db 3179 IPPPTIWLKNYKRIENSLEVRILSGGSKLOIARSOHSDSGNYTCIASNMEGKAKQY 3238
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360
Db 3239 FLISQVPPSVAGAEIPSDSVLGENVELVCNANGIPTPLIQLKDGKPIASGETERIRV 3298
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL----- 373
Db 3299 SANGSTLNIYALTSDTGKYTCVATNPAGEBRIFNLVNVVPTIRGNKDEAKLMTYVD 3358
QY 374 --K-----I-W-----SM-----KQ--D-----NCV----- 384
Db 3359 TSINIECRXTGTPPPQINLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVSN 3418
QY 385 -----H-----H-----D-----L--Q----- 388
Db 3419 RAGVDNKHYNLQVAFPPNDMSNGTEBITVLKSGSTSMACITDGTAPSMWLBDGQPLG 3478
QY 389 -----Q-----Q-----H-N-----KE 393
Db 3479 LDAHILTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKVLPEPPHINGSEHEE 3538
QY 394 I-----YT-----I--K--W-----N----- 399
Db 3539 ISVIVNPLELTCIASGIPAPKMTWMDGRPLPQTDQVOTLGGGEVLRISTAQVEDTGRY 3598
QY 400 -----SPTG-----P-----GT-----N----- 407
Db 3599 TCLASSPAGDDKEYLVRVHPVNPNTAGTDEPRDITVLRNRQVLTLECKSDAVPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS-A--S--F----- 420
Db 3659 NGERLOATPRVILSGRYLQINNADLGDNTANTCVASNIACKTREFILTVNVPNIK 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G-----I----- 433
Db 3719 GPQSLVILANKSVLECIAGVPTPRIITWRKDGAVLGNHARYSILENGLHIQSAHVTD 3778

QY 434 --C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHVPPIAPGPTNMTVIVVQTTLACEATGPKPSINW 3838
QY 440 -K--H-----Q-E-----P-V-----Y----- 446
Db 3839 RKNGLHLLNDQNSRYLLSSGSLVIPSVDATYECTVTNGAGDKRTVOLTVOVPP 3898
QY 447 SVA-----F-S-----P-D----- 453
Db 3899 STADEPTDELTKHAPAVITCTASGVPPFSIHWTKNIGIRLLPRGDGYRILSSGAIEILAT 3958
QY 454 -----GRY-----L--ASG--S-- 461
Db 3959 QLNHAGRYTCVARNAGSAHRHVTLHVHPPVQPOPSSELHVLNPNILLPCEATGTPSP 4018
QY 462 F-----D-----KCV----- 466
Db 4019 FITWQKEGINVNTSGRNHVLPSGGIQLISRAVREDAGTYMCVAQNPAGTALGKILNVQV 4078
QY 467 -----H-----I-W-----NT-Q----- 472
Db 4079 PPVISHLKEYVIADVDPITLSCADGLPPDITWHKDGRAIVESIRQVRLSSGSLQIAF 4138
QY 473 V-----C-----L-----HY----- 477
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRIRSTEGHYTVNENSOAILPCVADGPT 4198
QY 478 -----L-N-----GO-----VL-----L-----N----- 485
Db 4199 PAINMKDNVLLANLLGKYTABPYGELILENVVLEDSGFYTCVANNAGETHVSLTVH 4258
QY 486 -----L-G--R-SIC-----L-YT-----LP-H-----H--LV-- 500
Db 4259 VLPTFTPLPGDVSINKGEQLRLS-CKATGIPLKLTFWTNNNIIPAHFDSVNGHSELVIE 4317
QY 501 -----V-----I-PL-----V----- 505
Db 4318 RVSKEDSGTYVCTAENSQVFKATGVYVKEPPVFKGDPSPNWIEPLGNAILNCEVKGD 4377
QY 506 --AL-----IE-----L-----L-VL-----K 514
Db 4378 PPTTIQWNRKQVDIBISHRIQGLNGSLAIYGTVNE 4413

RESULT 13

Q8WZ42

ID Q8WZ42 PRELIMINARY; PRT: 34350 AA.
AC Q8WZ42;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DE 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Titin.
GN Name=TTN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309627; PubMed=10850961;
RA Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereuse F.,
RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,
RA Granzier H., Labeit S.;
RT "Series of exon-skipping events in the elastic spring region of titin
RT as the structural basis for myofibrillar elastic diversity.";
Circ. Res. 86:1114-1121(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21573839; PubMed=11717165;
RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,
RA McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,

RA Labeit S.;
RT "The complete gene sequence of titin, expression of an unusual ~700
RT kDa titin isoform and its interaction with obscurin identify a novel
RT Z-line to I-band linking system.";
RL Circ. Res. 89:1065-1072(2001).
RN [3]
RN SEQUENCE FROM N.A.
RA Centner S.B.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AJ277892; CAD12456.1; .
DR PDB; 1TNN; NMR; @=33480-33579.
DR PDB; 1TNN; NMR; @=33480-33579.
DR Genew; HGNC:12403; TTN.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004896; F:hematopoietin/interferon-class (D200-domain...; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0008307; F:structural constituent of muscle; IEA.
DR GO; GO:0007517; P:muscle development; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000282; Cytok receptor_2.
DR InterPro; IPR000577; FGGY_Kin.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR004168; PPAK motif.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr kinase.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR Pfam; PF00041; fn3; 132.
DR Pfam; PF00047; Ig; 112.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF02818; PPAK; 53.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00060; FN3; 132.
DR SMART; SM00408; IGC2; 65.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; UNKNOWN_1.
DR PROSITE; PS50853; FN3; 132.
DR PROSITE; PS50835; IG_LIKE; 143.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
SQ SEQUENCE 34350 AA; 3816189 MW; 5B120058A7CE58A CRC64;

Query Match 71.0%; Score 2592.4; DB 2; Length 34350;
Best Local Similarity 8.7%; Pred. No. 4.4e-40;
Matches 454; Conservative 50; Mismatches 8; Indels 4696; Gaps 388;

QY 1 M-----SIS-----SDE-----V 8
| | | | |
Db 4290 MIHTPLVDTVSBERGDIVHLTTSITNAKEVNWYFENKLVPSDEKFKCIQDQNTVTLVDKV 4349

QY 9 N----- 9
|
Db 4350 NTEHDQGEYVCEALNDGKTSATSAKLTIVKRAAPVIRKRIEPLVALGHLAKFTCIQSA 4409

QY 10 ---F---L---VY-----R---Y-----L---Q----- 17
| | | | |
Db 4410 PNVRFOWFKAGRIEYSDKSISSKVISSLEILRTQVDCGYTCASNEYSVSCAT 4469

QY 18 ---E---S-----G---F----- 22
| | | | |
Db 4470 LTVTEAVPTFLSRKSLTTFVGAAKFICTVGTPTVETIQKDGAAALSPSPNWKISDA 4529

QY 23 ---H-----S-----A-----F----- 26
| | | | |
Db 4530 ENKHILELSNLTIDRGVYSCASNRFGADICQAEIIIDKPHFIKELEPVQSAINKKVH 4589

QY 27 -----T-----F----- 28
Db 4590 LECQVDEDRKVTVTWSKQCKLPKPKDYKICREDKIATLIEIPLAKDKSGTYVCTASNEA 4649

QY 29 G-----IK-----S---H-----ISQ-----S 37
| | | | |
Db 4650 GSSSCSATVTVREPPSFVKVKVDPVSLMLPGESARLHCKLKGSPVI-QVTFWKNKSELSES 4708

QY 38 N-----IN-----G-----A-----LV---P-----PA- 47
| | | | |
Db 4709 NTVRMVFNSEAILDITDKVEDSGSYCEAVNDVSGSDCSTEIVIKPEPFIKTLEPAD 4768

QY 48 -----AL-----IS-----I----- 52
| | | | |
Db 4769 IVRGITNALLQCEVSGTGPFSEISWFKDKKQIRSKKYLRFSLVCLCLEIFSPNSADVGEY 4828

QY 53 -----I-----Q-----KG- 56
| | | | |
Db 4829 ECVVANEVKGCGCMATHLKEPPTFVKVVDLIALGGQTVTLQAAVRGSEPISVTWKQG 4888

QY 57 -----L-----Q-----Y-----VE 61
| | | | |
Db 4889 EVIREGKIKMFSNGVAVLIIIPDVQISFGKYTCIAENAGSQTSGBELIVKEPAKILE 4948

QY 62 -AE-----V-S-----I-----N-----E- 68
| | | | |
Db 4949 RAELIQVTAGDPATLEYTVAGTPELPKPKWKYKDGRLVASKKYRISFKNNVAQLKFYSDEL 5008

QY 69 -D-----G-----T-L-----F-----D-----G---R-----P 77
| | | | |
Db 5009 HDGQVTFEISNEVSGSSCETFTTDLDRDIAPFTFKPLRNVDSVNGTCLRLDCKIAGSLP 5068

QY 78 -----I-----E---SLSLI-----DA----- 86
| | | | |
Db 5069 MRVSMFKDGKEIAASDRYIAFVEGTASLEIRIVDMNDAGNFTCRATNSVGSKDSSGALI 5128

QY 87 -----V-MP---DV---V-----Q-----T-R----- 95
| | | | |
Db 5129 VQEPSPFVTKPGSKDVLPGSAVCLKSTFGSTPLTIRWFGKNKELVSGGSCYITKEALES 5188

QY 96 -----Q-----Q-----Q-A- 98
| | | | |
Db 5189 SLELVLVKTSDSGTCTCKVSNVAGVGECSANLFVKEPATFVEKLEPSQLLKKGATQLAC 5248

QY 99 -----Y-----R-----D---KL---AQ----- 105
| | | | |
Db 5249 KVTGPPPIKITWFANDREIKESKHKRMSFVESTAVLRLTDVGIEDSGEYWCBAQNEAGSD 5308

QY 106 ----- 105
| | | | |
Db 5309 HCSSIVIVKESPYFTKEPKPIEVLKEYDVMLLAEVAGTTPPEITWPKDNTILRSGRKYKT 5368

QY 106 --Q-----Q-AA-A-----A----- 111
| | | | |
Db 5369 FIQDHLVLSQILKFVAADAGEYQCRVTNEVSGSSICSAVTLREPPSFIKKIESTSLRG 5428

QY 112 -AA--A-----A-----A-----A- 116
| | | | |
Db 5429 TAAFOATLKSLPITVTWLKDSDEITDDNIRMTFENNVAISLYLSGIEVHKDGKYVCQAK 5488

QY 117 --A-----A-S-----Q-----Q---G-----SAK----- 125
| | | | |
Db 5489 NDAGIORCSALLSVKEPATITEANVSIDVTQGDPAVLQVFKSGTKEITAKWPKDGOELTL 5548

QY 126 -----N----- 126
| | | | |
Db 5549 GSKYKISVTDVTSILKIIISTEKKDSGEYTFEQNDVGRSSCKARINVLDLIIPSPFTKL 5608

QY 127 -----G-----E-----NTA----- 131
| | | | |
Db 5609 KKWDISKGFIDLECIAGSHFISIQWFKDDQEIASEKYKFSFHDNTAFLEISQLEGTD 5668

QY 132 ---N---G---E---N---G-A-----138
Db 5669 SGTTCATNKAHQCSGHLTVKEPPYFVEKPOSDVNPENTRVOLKALVGGTAPMTIKW 5728
QY 139 ---H---T-I-----A-----142
Db 5729 FXDNKELHGAARSVMKDDTSTSLFPAKATDSGTYYI COLSNDVGTATSKATLFVKEPP 5788
QY 143 ---N---N---N---H-----145
Db 5789 QFIKKPSPVLVRNGOSTTFECQITGTPKIRSVWYLDGNEITAIQKHGISFIDGLATFOI 5848
QY 146 ---T---D---MM-EV---D---G-153
Db 5849 SGARVENSCTYVCEARNAGTASCSTELVKVEPPTFIRELKPVEVVKYSDEVECEVTGT 5908
QY 154 ---D---V---E---I-----157
Db 5909 PPEVTWLNKNRIRSSKYYTLTDRVSFNLHITKCDPSDTGEYQICIVSNEGSCSCTR 5968
QY 158 ---PP---N---K---A-----V-163
Db 5969 VALKEPPSPFKKIENTTVLKSSATFQSTVAGSPPISTWLKDDQILDDEDDNVYISFVDS 6028
QY 164 V-L-R---GH---ES-----170
Db 6029 VATLQIRSVDNHSGRYTQAKNESGVERCYAPLLVOEPAQIVKAKSVDVTEKDPMTLE 6088
QY 171 ---V---E-----171
Db 6089 CVVAGTPELVKWLKDGKQIVPSRYFSMFENNVSFRIQSVMKDQSGQYTFKVENDFGS 6148
QY 172 ---V---F---I---C---A-W-----177
Db 6149 SSCDAVLRVLDQNIPPSFTKLTQMDKVLGSSIHMECKVSGSLPIAQAQWFKDQKEISTSA 6208
QY 178 ---N---P---V-----180
Db 6209 KYRLVCHERSVSLVNNLEEDTANYTCKVSNVAGDDACSGILTVKEPPSFLVKPGRQA 6268
QY 181 ---S---D---LL---AS-----186
Db 6269 IPDSTVEFKAILKGTGTPPKIKWFKDDVELVSGPKCFIGLEGSTSFNLNLSVDASKTGQYT 6328
QY 187 ---GS---GDST---A---R-IW--N-197
Db 6329 CHVTNDVGSCTTMLLVTEPPKFVKLEASKIVKAGDSRLECKTAGSPEIRVVWFRNE 6388
QY 198 ---L---S---E-N-ST--SG-----207
Db 6389 HELPASDKYRMTFIDSVAVIQMNNLSTEDSGDFICEAQNAGSTSCSTKVIIVKEPPVFS 6448
QY 208 ---QL---V---LR-----212
Db 6449 FPPIVETLKNAEVSLECELSGTGTPPEVWVWKQRQRSSKKYKIASKNFHTSIHILNVD 6508
QY 213 ---HC---V-----214
Db 6509 SDIGEYHCAQAEVSGDTCVTVLKEPPRFVSKLNSLTVVAGEPAELOASIEGAQPIFV 6568
QY 215 ---IRE---GG-----219
Db 6569 QWLKEKEEVIRESNIRITFVENVATLQFAKAEAPANAGKYICOIKNDGGMEENMATLMVL 6628
QY 220 ---V-PS-----N-----219
Db 6629 EBAVIVEKAGPMTVTGETCTLECKVAGTPELSVWYKDKLLTSSQKHKFSFYNKISSL 6688
QY 220 ---QD---V-PS-----N-----225
Db 6689 RILSVRODAGTYTFQVQNNVKSCTAVVDVSDRAVPFSTRRLKNVTGGVLGASCILEC 6748
QY 226 K---D-V-T---SLD-----232

Db 6749 KVAGSSPISVAMFHEKTKIVSGAKYQTPSDNVCTQLNLSLDSOMGNITYCVAANVAGSD 6808
QY 233 ---W-----233
Db 6809 ECRAVLTVQEPSPSFVKEPEPLVLPKGNVTFTSVIRGTPPFVKVMPRGARELVKGORCNI 6868
QY 234 ---N---S-E---G---T-LL---A---T-G-243
Db 6869 YFEDITVAELENFNIDISQSGEYTCVVSNAGQASCTTRLFVKPEAPFLKRLSDHSVEPGK 6928
QY 244 ---S-Y---DGF-----248
Db 6929 SIILESTYGTLPISVTWKDGFNITTSKCNIVTEKTCILEILNSTKRDAGQYSCIE 6988
QY 249 ---A-R---I---W---TK---254
Db 6989 NEAGRDVCGALVSTLEPPYFVTELEPLEAAVGSVSLOCQVAGTPEITVSWYKGDTKLRP 7048
QY 255 ---D-G---N-L--AS-----260
Db 7049 TPEYRTYTNVATLVFNKNVINDSGEYTCKAENSIGTASSKTVPFRIQERQLPPSPARQL 7108
QY 261 ---TL---G-----Q---H-265
Db 7109 KDIEQTVGLPVTLTCLNGSAPIQVCWYRDGVLLRDNENLOTFSVDNVATLKILQTDLSH 7168
QY 266 KG---P---I---F---A-LK-273
Db 7169 SQOYSCSASNPLGTASSARLTARBPKKSPFPDIKPVSIDVAGESADFECHVTGAQPMR 7228
QY 274 ---W---NK-----KG-----278
Db 7229 ITWSKDNKEIRPGGNYTTCVGNTPHLRLKLVKGDGSGQYTCQATNDVCKDMCSAQLSVK 7288
QY 279 ---N---FI-----281
Db 7289 BPPKFVKLEASKVAKQGESIQLECKISGSPBKVSFWRNDSSELHESWKYNNMSFINSVAL 7348
QY 282 L---SA---GV-D---K---T-----289
Db 7349 LTINEASAEADSGDYICEAHNGVGDASCSTALTVAAPPVFTQKPSVPGALKGSDVILQCEI 7408
QY 290 ---T---IIV---D---A---H-----296
Db 7409 SGTPEVWVWKDRQVRNSKKFKITSKHFDTSLHLNLEASDVGEYHCKATNEVGS DTC 7468
QY 297 ---T-G--E-A---K-----301
Db 7469 SCSVKFKEPPRFVKLSDTSTLIGDAVELRAIVEGFQPISVVWLKORGEVIRESENTRIS 7528
QY 302 ---Q-----Q-----303
Db 7529 FIDNIATLQGSPEASNSGKYICOIKNDAGMECSAVLTVLEPARIIEKPEMTVTGTNP 7588
QY 304 F---P---F---H-S--A---P-A---L-----312
Db 7589 FALECVVCTGTPBLSAKWFKDGRSADSXKHITTFINKVASLKI PCAEKSDKGLYSPEVK 7648
QY 313 ---D-V---D-----W-----316
Db 7649 SVGKSNCTVSVHVSDBRIVPPSFIRKLKDVNAILGASVLECRVSGSAPISVGVFQDGN 7708
QY 317 ---QS---N-----N-----320
Db 7709 VSGPKQCSFSBNVCTLNLSLEPSDTGIYTCVAANVAGSDSCSAVLTVQEPSPFQTPD 7768
QY 321 ---TP-----282
Db 7769 SVEVLPGHSLTFTSVIRGTPPFVKWFKGSRRLVPGESCNI SLEDFTVLELELFEVQPLES 7828
QY 323 ---ASC-----S-----T-----327

Db 7829 GDYSLVNDAGSASCTTHLFVKEPATFVKRLADFSVETGSPVILEATYTGTPPSVSWI 7888
QY 328 -D-----M-----C-----330
Db 7889 KDEYLISQSERCSITWTEKSTILBILSTIEDYVAQYSLIENAGQDICEALVSVLEPPY 7948
QY 331 -I-----HV-----CKL-G-----Q-----338
Db 7949 FIEPLEHVEAIGEPATLQCKVDGTPAIRISWYKEHTKLSAPAYKQFKNNVASLVINK 8008
QY 339 -D-----D-----R-P-----I-KT-----F-----345
Db 8009 VDHSVDGEYCKADNSGVAVASAVLVIKERKLPFPFARKLDVHETLGLFPVAFECRNG 8068
QY 346 -Q-----G-----H-----T-----N-----350
Db 8069 SEPLQVSWYKDGVLKDDANLQTSFVHNVAQLQLQDQSHIGQYNCASNPILGTASSA 8128
QY 351 -EV-----EV-----NA-IK--W--D-----PTGN-----362
Db 8129 KLILSEHEVPPFPDLKPVSDALGESGTFKCHVTGTAPIKITWAKDNREIRPGNYKWT 8188
QY 363 -L-L-----A-SC-S-----D-----DM 371
Db 8189 LVNTATLTVLKVGKGAGQYTCVASNIAGKSCSAHLGVQEPFRFIKLEPSRIVKQDE 8248
QY 372 -T-L--KI-----W-----SMK-Q-----D-----C--383
Db 8249 FTRYECKIGSPEIKVLWYKDET-EIQESSKFRMSFVDSVAVLEMHNLNLSVEDSGDYTCEA 8307
QY 384 -V-----V-----H-----D-----LQ-----Q 389
Db 8308 HNAAGSASSSTSLKVKPEPIFRKKPHIETLKGADVHLECELQGTTPPHVSWYKDKREL 8367
QY 390 -H-----H-----N-----N-----391
Db 8368 SGKKYKIMSENFILTSIHILNVAADIGEYOKATNDVGSCTCVGSIALKAPPRFVKKLS 8427
QY 392 -KEI-Y--TI-----K-----W-----399
Db 8428 ISTVVGKEVQLQTTIEGAEPISVVMFKDGEIVRESNHWISYSENIATLQFGRVSPAN 8487
QY 400 -S-----S-----P-----TG-----P-----404
Db 8488 GKTCQIKNDAGMQECFATLSVLEPATIVKEPKSIKVTGDTCTLECTVAGTSELSTKWF 8547
QY 405 -G--T-----N-----N--P-----N-----A-----411
Db 8548 KDGKELTSNDKYNKISFFNKVSGLKIINVPASDSGVYSFEVQNPVKGDSCTASLQVSDRTV 8607
QY 412 -N-L-----M-----L-----ASAS-F-----420
Db 8608 PPSFTRKLKETNGLSGSVVMCKYGVSPISVSWFHEGNEISSGRKYQTLTDNTCAL 8667
QY 421 -DS-----DS-----TVR-----L-----426
Db 8668 VNMLESDSGDYTCIATNAGSDECSAPLTVREPPSVQKPDMDVLGTGNTVFTSIVKG 8727
QY 427 -W-----W-----DVD--R-G--ICI-----435
Db 8728 TPFPSVSWFKSGSELVPGDRCNVSLSDSVALELFDVDTQSGEYTCIVSNEAGKASCTT 8787
QY 436 H-----H-----TLT-----K-----H-Q-----442
Db 8788 HLYIKAPAKVVKLNYSIEKGPLILEGTFPTTPISTVTKNGKNGINVTSPQRCNITTTTE 8847
QY 443 -E-P--V-----YS--V-----A-----F-----SP-----D-----453
Db 8848 KSAILEIPSTVEDAGQYNCYNIENASGKSCSAQILILEPPYFVKQLPEPKVSGVDSASL 8907
QY 454 -G-----G-----R-----Y-----L-----A--S-G 460
Db 8908 QCQAGTPEIGVSWYKGDTKLRPTTTYKQHFRRNNVATLVFNQVDINDSGEYICKAENSVC 8967

QY 461 -----SF-----D-K-----C-V-----HI--W-----469
Db 8968 EVSASTFLTVQEOKLPSPFSRQLRDVQETVGLPVVDFCAISGSEPSVSWYKDGKPLKDS 9027
QY 470 -----NT-----QV-C-----L-475
Db 9028 PNVQTSFLDNTATLNIKFKDRSLAGQVSCATNPIGSASSARLLILTEGKNPPFFDIRLA 9087
QY 476 -----H-----Y-L-N-----479
Db 9088 PVDVVGESADFECHVTGTQPIKVSNAKDSREIRSGGKYQISYLENSAHLTVLKVDKGDS 9147
QY 480 GO-----V-----L-----L-----N--L-GR--S-----489
Db 9148 GQYTCYAVNEVGKDSCTAQINIKERLIPSFYTKRLSETVEETEGNSFKLEGRVAGSQPIT 9207
QY 490 -----I-----C-----LYT-----L-----495
Db 9208 VAWYKNNIEIQPTSNCEITFKNTLVLQVRKAGMNDAGLYTCKVSNDSAGSALCTSSIVIK 9267
QY 496 -----P-----H-----H-----498
Db 9268 EPKPKPEVFDQHLTPVTVSEGYVQLSCHVQSGSEPIRIOWLKAGREIKPSDRCSFASGT 9327
QY 499 -----L-----L-----L-----499
Db 9328 AVLELRDAKADSGDYVCKASNVAGSDTTKSKVTIKDKFAVAPATKAAVDRGLFFVSEP 9387
QY 500 -----VV-----IP-----503
Db 9388 QSIIRVVEKTTATFIKVGDPPIPNVKWTKGKWRQLNQGRVFIHQKGDDEAKLEIRDTTK 9447
QY 504 -----L-----VA-----L-----L-----L-----V-LK 514
Db 9448 DSGLYRCVAFNEHGEIESNVNLQVDERKQEKIEGDLRAMLKKTPIK 9495

RESULT 14

Q6XA09 PRELIMINARY; PRT; 7191 AA.
AC Q6XA09
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Nonribosomal peptide synthase.
GN Name=NRPS1;
OS Alternaria brassicae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Pleosporaceae; Pleosporaceae; Alternaria.
OX NCBI_TaxID=29911;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14727058;
RA Guillemette T., Sellam A., Simoneau P.;
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria
brassicae and flanking genomic sequences.";
RL Curr. Genet. 45:214-224(2004).
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC EMBL; AY246697; AAP78735.1; -;
DR InterPro; IPR010071; AA_adenyl_dom.
DR InterPro; IPR009081; ACP_like.
DR InterPro; IPR000873; AMP_bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR006163; Pp_bind.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; Condensation; 9.
DR Pfam; PF00550; PP-binding; 6.
DR PRINTS; PRO0154; AMPBINDING.
DR TIGRPFAM; TIGR01733; AA-adenyl-dom; 4.

213 QY -----H-----C-----I-- 215
3561 DB AONMAETFKILDIVVCSASELTIGKLDILTESSIHQIMENWNPOLPPPIRCLHVDH 3620
216 QY -----R-----EG--G--QD-----VP----- 223
3621 DB QALTRPRTTKAVEGWDGTFYQDPDKITNOQAVHLSQIGVTTFTFVILPEKSYAIVSM 3680
224 QY ----- 223
3681 DB IAIMKAGAVVPLDPKHPQTRLELIEDVGASVVLCSRGYHTTASEVAKTAVIVDQSR 3740
224 QY -----S-----N-----K----- 226
3741 DB KLGVPISSPRTCATPDNAYCULFTSGTCKPKCTIIPHQAFCSTAAAFTRMNINATSR 3800
227 QY -----D-----VTS--L-----D----- 232
3801 DB TFQFASYTFDASCIEILSALTUGATVCVPTEDDRMNAAGAIRKLRVNMELLTPSVLGTI 3860
233 QY -----W--NS-----E----- 236
3861 DB EPERVGLKTVSGEALSGPILKKWSNSTCFINAYGPTSCSVAAATAYKSTLDHKLIVS 3920
237 QY --GTLATGS--Y-----D-----G--FAR-----I-- 251
3921 DB EPCT--IGTSGCRLWIVHPRNHDKLMVPSVGVGLVIEGPTVARGYLNDVKTAKAFINDP 3979
252 QY -----T-----K-----DG--NL--AST--L--GO-----H-- 265
3980 DB AWAKTIFSSNNTFEAARMYKTDLVRYNTDGSVNYIGRKDTQIKLANGQRIELGEIFHV 4039
266 QY ----- 265
4040 DB KNPPERVQSAVELVAPNSRSSAKALAVFAVQDQIDGEOQSVQPASTDLPAAADLLPL 4099
266 QY -----K-----G-----P--IF--ALKW----- 274
4100 DB SDELDMCKNTENGLAGSLPSYMPIAIFIPVTKLPWTSAGKLDNRNLSIVQNLSTWMA 4159
275 QY -----NKK----- 278
4160 DB MYRLTSIANKKPKITEAEKIKHKAVCVSLPSPSSVGIDDSFVRLGGDSISSMRLVAMAH 4219
279 QY -----N-----P-----I----- 281
4220 DB TEQWELSFDFIKNPKLSDAKIGAQISKSQAQKVMQPPDLLPASLITRSDVISEVVQOC 4279
282 QY -----LS--AG--V-----DKT----- 289
4280 DB QVSKEDLDQAYPTSSLDALLTSLIKQAGAYVAQHVLAALPKSLDMTKPKAAWESAIOEID 4339
290 QY -----T--II-----W-----DA-----H-----T-- 297
4340 DB ILTRTIQMPGIFMOAVLRNDPVDWREAKSLKSAEDDASKIPPHLGHLAAVTLVTPTS 4399
298 QY -----A-----K-----Q----- 302
4400 DB GERYFVWTLHLYDGSYILMLQVQOQIYKGVSTTPQTSYARFVEYLSSTSVSDSDVY 4459
303 QY -----QPP-----F--HS--A-----PA----- 311
4460 DB WRERLTGVNAYQFPRPSHATSSAPPNGQMPQHSMKIAHRKNTDVTTPANAIKAAWALLAA 4519
312 QY -----L-----DV----- 314
4520 DB YTGSDVDVFGETLAGRDVAMTGITDVCGPTLTTPVSRVKIDRGATVSDLLNTIATNIDR 4579
315 QY -----D--W-----Q----- 317
4580 DB IPYQHGLSAIKALGEDMIAACDFQNLVLOTENEELADSMWSVHDNEQGNFTYPLVI 4639
318 QY -----S-----N----- 319

4640 DB ECKWGLSKTEVLAFDANVISLWHVORLIYQFETVLIQOSATHVRIHAVLSQDQQLVR 4699
320 QY --N-----T-----F-----AS----- 324
4700 DB KKNAYEPRLIDDTVPSLFFKKAASQPTTTAVTAFDGEFSYVGSALASOLAQELVKFGAG 4759
325 QY -----C--S-----TD----- 328
4760 DB PECLIPICVDKGRMAIVAILAILISGAGVYVPLSPDTPASRHLHIVETCKASIVLCSPKYT 4819
329 QY -----N-----C-----I--H-- 332
4820 DB HRFVEMVGHVPSVSETAIRQLPTSSISLSQRAKNNICVIFTSGTGTLPGKGVVIEHXS 4879
333 QY -----VKLG-----Q-----D-----R-----P----- 341
4880 DB VSSAAIACE--GLHITPTSRVQFCSFLFDVSVGETLTVLLRGATICVPSDEQRTTNLAA 4938
342 QY IKT-----F-----QG-----HTN----- 350
4939 DB V-TDLNANWAFPTPSVASTLEGPKSVPTTLETLVGGEAMTSDVVDKMATGVNLH--NGYGP 4996
351 QY --E--V--N--A--I--K--W--D--P--T--G-- 361
4997 DB TEGTVPAGNDHVSAQORDPSNIGHPLKSGRAMLTNSDNPHELAPIGATAELCLEGLLAR 5056
362 QY -----N-----L--LA--S----- 366
5057 DB GYLNDPKRTSEAFIEAPAFKFNFSNSESRIYRTGDLVQYAADGSIQYMGKRDNIQKLAG 5116
367 QY -----C-----SD----- 369
5117 DB QRIELDEIOAVHANNVHVQVVPKVGCPCKKLTVVVSPFGTAASNAGSWRRILSDT 5176
370 QY -----D-----M-----TL-----K--I--W--SM--K--Q-- 380
5177 DB ESLSQINRARDRLADLVPSYVMPFIWIAVPRIPTILASAKLDKKQVGLWLEGMDEALYQRI 5236
381 QY -----D-----N-----C----- 383
5237 DB MGAELPEDMEGPGAAALTVLRGIWAKVLRNRPVEDVKPSKWSLSLGGDSISAMKLLAKCRS 5296
384 QY -----V-----H--D--L--O-----Q-----H 390
5297 DB EGINLNLQILRAKSLSHLAADVKSVMILDEKEQNDREPFALSPIORFVFEAGSIENSTH 5356
391 QY -NKE-----I--Y-----TI-----K-----W-----SP----- 401
5357 DB FN-QSSTLRILRYVQPMVQORALNSIVECHSMRLARFSDKNNNGQWQOLVMSKVSGSYAFT 5415
402 QY -----TGP----- 404
5416 DB AHDVSTASAAGAIISSTOKSLDIRTPGFVAVDLFNLKGHQMFLPMAHHLVIDVWSWGILL 5475
405 QY -----G-----T--N-----N----- 408
5476 DB GLEDLLGSGPVTTLPRSLPFTWCNMQTSSASEITQOLTVMKQPLVVEPAPFPMGMDV 5535
409 QY -PN-----A--N-----L-----MLASA--SFDS-----TV 424
5536 DB RPNVYGDSEDRDEFIDKETSAMAFDNHVVKTDLVDLILA--AILHSP--SRVFINKAPTL 5593
425 QY -----R--LWD--VD--R--G-----I-----C--IHTL-----T--K----- 440
5594 DB FNESHGREVMDGSLDLSTRTVGWFTTLPVPTVPIDDDEDEVIHTLRQVKDTRRKVASNGR 5653
441 QY -----HO-----E-----P--V--Y-----SVAFSP-----D-- 453
5654 DB PYFAHYLTEDGKERFANHAPMEVLFNVLGRQEQSHSDSL--LSPQVEGDDDETSDVGV 5712
454 QY -----G-----RY-----LA----- 458

Db 5713 KTSRMALFELASVTEGQIQLSFMNRYNRYKNOKGIRRTWABQORTLEETIAELAKTRDQ 5772

QY 459 -----S-----G-SFD-----KC-V--HI-----W 469

Db 5773 PTWADPPLPLBSYSRLDRVLKTLPHAGVPSFDQVEDMYPCSPIQDMILS QIKSPESYW 5832

QY 470 --NT-----QV-----473

Db 5833 SSTTEVRSKRGFVDASKVVDGKQVVARHPALRTIFDSVCKGVFDQIVVKSPDSGIV 5892

QY 474 --C-----L--H-YLNG-----Q-----VLN-----485

Db 5893 TYKDAELATLESIRHSSSLGKKKPVLPHQAAVVQTSKGKIFVKVIVNHAVIDGGSIG 5952

QY 486 --L-----GR-----S-----I-----C-----491

Db 5953 VIGQDLQEAEGRLSDGELYSAYIKYLRALPAEDAIAWKALRGVSCYFPTTPRDP 6012

QY 492 -----L-----L-YT-----494

Db 6013 KPRQLQSDMRFRFDELHDLAESSNVTTIANLLAAWALVLRSYTNSDVCYGLTSGRN 6072

QY 495 -----I-----PH-----H-----498

Db 6073 VPIDRIENAVGAFINMLVSRIELKSATSLLEIVENVQSDVFVGSMPHQHCSLAQFQHDGL 6132

QY 499 --L--VV-I--P--L--V-----AL--I-----508

Db 6133 SKSLFNTAVSIQRRSPELSDSGIEFQDLGHDHPSEPAITVVIDATRDGVRFTY 6192

QY 509 -----E-----L-----VL-----K 514

Db 6193 WSDAVTGDGEAKNVSTLMKILVQLSNPK 6221

RESULT 15

AAP78735 PRELIMINARY; PRT; 7191 AA.

ID AAP78735

AC AAP78735;

DT 01-APR-2004 (TREMELrel. 27, Created)

DT 01-APR-2004 (TREMELrel. 27, Last sequence update)

DT 01-APR-2004 (TREMELrel. 27, Last annotation update)

DE Nonribosomal peptide synthase.

GN NRPS1.

OS Alternaria brassicae.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;

OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.

OX NCBI_TaxID=29911;

RN [1]

RP SEQUENCE FROM N.A.

RX PubMed=14727058;

RA Guillemette T., Sella A., Simoneau P.;

RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria brassicae and flanking genomic sequences.";

RL Curr. Genet. 45:214-224(2004).

DR EMBL; AY246697; AAP78735.1; .

SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.0%; Score 2590.7; DB 2; Length 7191;

Best Local Similarity 9.4%; Pred. No. 2.2e-41;

Matches 442; Conservative 55; Mismatches 9; Indels 4203; Gaps 363;

QY 1 M-----S-----IS-----S-----D-----E-----7

Db 1521 MAFRLAGSKGRAPAGHAEVTLARLWASVLKLDAGAVGEDSPFRIGGDSISAMKLVTAR 1580

QY 8 ---V-N---F---L---V-----Y---R-----14

Db 1581 KGVVLNVASVFAQPKLLAMAATAVMLPSDDAKPEADTLPMELLIPASSRQAIVALAASE 1640

QY 15 -----Y-----LQ-----18

Db 1641 CDVFPDCIEDMYPCSKLQBLVMTLNKDPGTVVQPIYRLPADMDISRPKCAWKVVVAE 1700

QY 19 -----S-----GF-----S-----H-----S-----A-FT-----27

Db 1701 ASLRTRIVYSEHGFQVVVREDIKWQSLDPIQHINETTRRLPAKNGAPLATFTLVGENT 1760

QY 28 ---F-----G-----IK-----S-----H-----IS-----QS---37

Db 1761 DSMFFVWTAHHAHYDGMWSTALFRKVENYRGVHSMPTATVPYSRFFVKYISSLDQSDA 1820

QY 38 -----NI-----NGAL-----V-----P-----P-----AA-----L---49

Db 1821 FWLTQLDNVTAAQFPQLSPDRHVEANGOLLHTVLLTRNPGLEVTVPSMIRAANGILLAT 1880

QY 50 -----I-SII-----55

Db 1881 YSGSDDIVGETNSGREASVPGIESIIGPTITTAAPVRLNRSLTVHDYLKETQKQSSLS 1940

QY 56 ---GLQ-----Y-VE-----AE-----VS-----I-----66

Db 1941 LPYQFAGLQHISKLSPETAIACDFOSFLGIEAGDDFDAESSLMNMVMSANTIGTDFFSYAF 2000

QY 67 ---N-----E-----D-----GT-----71

Db 2001 VFNCKVATGVQVEALFDDRVRVQTLAQRMVQOQFILTQLNGTDMILRSLDLDLINPA 2060

QY 72 -----L-----F-----D-----74

Db 2061 DRKTISSMNSKVPPIIPRCIHSVIAEDQTALPTACAIADAWDTGFMYSRELDERSALAH 2120

QY 75 ---G-RP-----78

Db 2121 QLIRLGVRPKQFVPLCFDKSGMTIVAIIVAKAGAAFPVLPFEAPVLRUREIVSDIDADL 2180

QY 79 ---E-----S-----L-SL-----83

Db 2181 LLCAPOYRELQOSIPCSTWVVDROATETVAGRLPSLPSVHSDSPAYAFYTSGSTGPKGA 2240

QY 84 -----I-----DA-----V-----M-----PD-----90

Db 2241 VVHHTHWTSSTPAAGWKISTASRVLPASVYTFDACLIEVSTLMQGGTVCVPDQGSRT 2300

QY 91 -----VV-----Q-----93

Db 2301 NDLVGVINRFNVNWAALTSPSVVRMIVPSEIPQLETFLFVGEAMSQQDLVTWADKVNLGNG 2360

QY 94 -----T-----R-----95

Db 2361 YGPTCAAVATSNIMTPHMRPNNLGRAVTARGWIVSRNNHHTLAPVGAIGELLLEGAVG 2420

QY 96 ---Q-----Q-----Y-R-D---101

Db 2421 AGYLNAPKTAQVFGQARWCVGLMGDDDISAPVRYKTGDLVKYNEGTMLYLGRKDLQT 2480

QY 102 -----KL-----A-----Q-----Q---106

Db 2481 KVRGORLELSEVEHKLLDDHMQSALASVPTTGPCKAKRLVAIVSLQKHGDHDTDTDKQLR 2540

QY 107 ---Q-A---A-----A-----A-----A-----A-----113

Db 2541 LLPOENASLINIATIRDCGLERLPAYMIFSLWIAVERFPLMPSGKMDRRCAIQWLEQMDQA 2600

QY 114 ---A-----A-----A-----AA---118

Db 2601 TYRLISAMGTTDAEVNNGSPIEKKLRAPAKVNLNSVADVRLNQSFHLHGSDSIAMQVS 2660

QY 119 SQ---QG-----122

Db 2661 SQCRAQGPISVDIIRKSISAMASAVDLSQSSQVTTTEAKDYDLPLDPLSPIQKVFFDAV 2720

QY 123 -----SA-----KN-G-----E-----128

Db 2721 GDKHNYFNQTEFLRLSRNIETBELRSALTVLAKTHPMLRARFSKNEAGAKORIEKDVSS 2780

QY 129 -----NTAN-----G-----E----- 134
Db 2781 SFLRHHVQAGNDANLRPIIDHSQATLDVAKGPTFAIELFDVDDTFPSQAIALVAHLII 2840
QY 135 -----EN-----G--A-----H-----T-----IA----- 142
Db 2841 DVVSWGILLEDLQGLQLOPAPQSMYPYHMLHEQSLQATQESARRVFPVGDIPAGDLDY 2900
QY 143 -----N--N-----H--T--D--MM-----E-----V--D----- 152
Db 2901 WAMEGRPNVNGDVEEDLHLSTRDTWLLGAQDALATETIDILIAALFESFRKVPFDRST 2960
QY 153 -----G-----D-----V-----E-----I----- 157
Db 2961 IAHNEGHEGRETNNRQDLSRTVGFSTVTPHLPVPLNEATDMISTIRWRDFRNRTPD 3020
QY 158 -----P-----P-----NK----- 161
Db 3021 KGRPYFAYRNLTEEGOTRFASHWPAEAVFNYVGRLOQDNKDGCLFTALSVDVSREVGEDV 3080
QY 162 -----C--A-----W-----AV----- 164
Db 3081 PRLAFDITAAVSOQAISLFGWNRNRQKEIRAWVGKRCOTLVDVAVELLQARQERSV 3140
QY 165 -----L-----R-----G-----H--E----- 169
Db 3141 GNFYLPLLYNGISRLSAILPAGINLNDVEDIYPASPMQOGLLHTQSRPELYTYHTVSQ 3200
QY 170 --S-----E-----VFI----- 174
Db 3201 VQSADGNPIDPRLAEAWQVVHHRHQALRTIFIDSLAKDQKQIVLKEKPGRVQILADC 3260
QY 175 -----C--A-----W-----N----- 178
Db 3261 DDSQVANLLRHQSSIDCREALPPHRMSICKTKTGRVWFKLELSHVINDGTSVSNLLADLA 3320
QY 179 -----P-----V-----S--DL--L--A--SG----- 188
Db 3321 RAYARKLTRADAGPLYSDYIGYMLSRSSDADLAYWAHLSGIEPCLFPVLNDGIPSSPES 3380
QY 189 G--D-----START-----W-----N----- 197
Db 3381 GSVDELSTSRVQDFCKQNGVTLNSVLQTLWALTLYHYVGTDFVSFGLIASGRDIPVTN 3440
QY 198 LSE-----N-----S--T-----SGST-----OL----- 209
Db 3441 INEAVGCFVMVMSRLSFSDETTIAQLLEALQGSTALSHQCSLADIQHALQLPSLFN 3500
QY 210 -----L-----R----- 212
Db 3501 TATFORSLSPEDTALIYEDMEADAGEYIVTVNADVTQOSITVDFGYSKDRILPSQ 3560
QY 213 -----H-----C-----I----- 215
Db 3561 AQNAETFKLDSIVVCSASELTIGKLDILTESSIHQIMWNPQLPPPIRRCLHDVHD 3620
QY 216 -----R-----G-----QD-----VP----- 223
Db 3621 QALTRPRTTKAVERGWDGTFYQDFDKITNQLAVHQSIGVTTETFPILFEKSSYAIVSM 3680
QY 224 ----- 223
Db 3681 IAIMKAGGAVLPDPKHPQRLRELIEDVGASVVLCSRGVHTTASEVAKTAVIVDORSIR 3740
QY 224 -----S-----N-----K----- 226
Db 3741 KLGVPISSEKPTCATPDNAAVCLFTSGTKPKGTIIPHQAFCTSAAAFRRMNINATSR 3800
QY 227 -----D-----VTS--L-----D----- 232
Db 3801 TFOFASYTFDASCIEILSALTVGATVCVPTEDDRMNNAAGAIRKLNRWMSLLTPPSVLGTI 3860
QY 233 -----W--NS-----E----- 236

Db 3861 EPERVPLKTLVSGGEALSGPILKKWSNSTCFINAYGPTSCSVAAATAYKSTLDHKLIVS 3920
QY 237 --GTLIATQS--Y-----D-----G--FAR-----I----- 251
Db 3921 EPGT-IGTSGCRLWIVHPRNHDKLPVGSVGVGLVIEGTPTVARGYLNDVKTAKAFINDP 3979
QY 252 -W-----T-----K-----DG--NL--AST--L--GO-----H-- 265
Db 3980 AWAKTIFSSNNTFEARMYKTGDLVRYNTDGSVNYIGRKDTQIKLNGQRIELGEIEFHV 4039
QY 266 ----- 265
Db 4040 KNFPERVQSAVELVAPSNRESSAKALAVFAVVDQDAIDGEOQSVQAPASTDLPAADLLPL 4099
QY 266 -----K-----G-----P--IF--ALKW----- 274
Db 4100 SDELDMCKNTENGLAGSLPSYMIPIAIFIPVTKLPMTSAGKLDNRNRLSLVQNLSETWA 4159
QY 275 -----NKK-----G----- 278
Db 4160 MYRLTSIANKKPIITEAEKKIHKAVCSVLSPSSVGIDDSFVRLGGDSISSMRLVAMAH 4219
QY 279 -----N-----F-----I----- 281
Db 4220 TEOMELSFIDIENPKLSDLAKIGAKISKSQAQKVMQFPDLLPASLTRSDVISEVVQOC 4279
QY 282 -----LS--AG--V-----DKT----- 289
Db 4280 QVSKEDQDAYPTSSIQDALLTSLIKQAGYVAQVHALPKSLDMTKFKAAMESAIOEID 4339
QY 290 --T--II-----W-----DA-----H-----T-- 297
Db 4340 ILRTRIOMPSGIFMQAVLRNDPVDWREAKSLKSAEDDASKIPPHLGHLAAYTLTTPS 4399
QY 298 GE-----A-----K-----Q----- 302
Db 4400 GERYFWTLLHALYDQMSIYLMQRVQOIQYISKGVSTTPQTSYARFVEYLSSTSVDSDVY 4459
QY 303 -----QFP-----F--HS--A-----PA----- 311
Db 4460 WRERLTGVNAYQFPRPSHATSSAPPNGOMFOHSMKIAHRKNTDVTAPANAIRAAWALILAA 4519
QY 312 --L-----L-----DV----- 314
Db 4520 YTGSDDVVGETLAGRDVAMTGTIVCGFTLTVPSRVKIDRGATVSDLLNTIATNIDR 4579
QY 315 -----D--W-----Q----- 317
Db 4580 IPYQHGLSAIKALGEDMIAACDFQNLVLIQTENEELADSMWSVHDNEEQGNFTYPLVI 4639
QY 318 -----S-----N----- 319
Db 4640 ECKWGLSKTEVLAHFANDVVISLHWQRLIYQFETVLIQSQATHVRHIAVLSDQDKQLVR 4699
QY 320 --N-----T-----F-----AS----- 324
Db 4700 KMWAYEPRLIDTVPPLFFPKKAASOPTTAVTAFDGEFSYGELSALASQLAQLVFGAG 4759
QY 325 -----C--S-----TD----- 328
Db 4760 PECLIPICVDKSRWAIVAILAILISGAGVYPLSPDTPASRHLHIVETCKASIVLCSPKYT 4819
QY 329 -----M-----C-----I--H----- 332
Db 4820 HRFVEMVGHVFSVETAIRQLPTSSISLSQRAKNNICVIFTSGSTGLPKGVIEHKSI 4879
QY 333 -----VCKLG-----Q-----D-----R-----P----- 341
Db 4880 VSSAAICE-GLHITPSTRVFCFLFVSGVETLTVLLRGATICVPSDEQRTNLAAA 4938
QY 342 IKT-----F-----QG-----HTN----- 350

Db 4939 V-TDLNANWAFLETPSVASTLEGPKSVPTLETTLVVGGEAMTSDVVDKWTGVNLH-NGYGP 4996
QY 351 -E-V-N-A-I-K-W-D-P-T-G- 361
Db 4997 TEGTVFAIGNDHVSAGQDPDSNCHPLKSGRAWLTSNDNPHELAPICATAELCLEGPLLAR 5056
QY 362 -N-LA-S- 366
Db 5057 GYLNDPKRTSEAFIEAPAFKPNFSSNESRIYRTGDLVQYAADGSIQYMGKDNQIKLAG 5116
QY 367 -C-SD- 369
Db 5117 QRIELDEIOAVHADNNHVQVQLPKVGFCTKLTVVVSPFGTAASNAGSDWRILSDT 5176
QY 370 -D-M-TL-K-I-W-SM-K-Q- 380
Db 5177 ESLSQINRRARDLADLVPSVMVFFIWIAPRPTLASAKLDKKQVGLWLEGHDEALYQRI 5236
QY 381 -D-N- 383
Db 5237 MGAELPEDMEGPGAAALTVLIRGIWAKVLRNPVEDVKPSKWSLSLGGDSISAMKLLAKCRS 5296
QY 384 -V-H-D-L-Q- 390
Db 5297 EGINLNLQILRAKSLSHLAADVKSVIDLHEKQBNDRPALSPIORFYVEAGSIENSTH 5356
QY 391 -NKE-I-Y-TI-K-W-SP- 401
Db 5357 FN-QSSTILRIVQVQAMVORALNSIVECHSMLRARFSDKNNGQWQQLVMSKVSQSYAFT 5415
QY 402 -TCP- 404
Db 5416 AHDVSTASAAGAIISSTOKSLDIRTGPFAVDLFLNLKGHQLFMAAHVLVIDVWSGILL 5475
QY 405 -G-T-N- 408
Db 5476 GLEDLLGSGPVTTLPRLSPFQWCKMOTSSASEITQUTVKNQPLVVEPANFAPFWMGMDV 5535
QY 409 -PN-A-N-L-MLASA-SFDS-TV 424
Db 5536 RPNVYGDSEDFVIDKETSAFNDNHVYKTDLDVILIA-ALLHSF-SRVFINRKAFTL 5593
QY 425 -R-LWD-VD-R-G-I-C-IHTL-T-K- 440
Db 5594 FNESHGREVWGSNLDLSRTVGWFTTLYPVTVPIDDEDEVIHTLRQVKDTRKVASNGR 5653
QY 441 -HQ-E-P-V-Y-SVAFSP-D- 453
Db 5654 PYFAHRYLTEDGKERFANHAPMEVLFNLYLGRQESGHSLSL-LSPTQVREGDDDETSDBGV 5712
QY 454 -G-RY-LA- 458
Db 5713 KTSRMALFEISAVTEGQIQLSFMYNRYSKNQKIRRTABQORTLEETAEIABLAKTDRPQ 5772
QY 459 -S-S-SFD-KC-V-HI-W 469
Db 5773 PTWADPPLPLESYSRLDRVLKTLPHAGVSPFQVEDMYPCSPIQDGMILSQQKSPESYW 5832
QY 470 -NT-QV- 473
Db 5833 SSTFEVRSKRGPDVADSKVVDGWKQVVARHPALRTIFIDSVCKGVFDQIVVKSPPDSGIV 5892
QY 474 -C-L-H-YLNG-Q-VLLN- 485
Db 5893 TYKCADELATLLESIRHSSSLGKKKPVLPQAAVQVQTSSTSGKIFVKVIVNHAVIDGGSLG 5952
QY 486 -L-GR-S-I-C- 491
Db 5953 VIGQDLQEAVEGRLSEDGPLYSAYIKYLRALPAEDAIAWKAKLRCVSPCYFTPTPRDPS 6012
QY 492 -L-YT- 494
Db 6013 KPRQLQSLDMRFRFDELHDLAESSNVTTIANILLAAWALVLRSYTNSSDVCYGLTSGRN 6072

QY 495 -----L-----PH-----H----- 498
Db 6073 VPIDRIENAVGAFINMLVSRIELKSATSLSLEIVENVQSDFVGSMPHQHCSLAQFQHDIGL 6132
QY 499 -----L-VV-I-P-L-V-----AL-I----- 508
Db 6133 SGKSLFNTAVSIQRRSSPEELTSDSGIEFQLDGHDPSFAITVNIIDATRDDEGVRFY 6192
QY 509 -----E-----L-----VL--K 514
Db 6193 WSDAVTDGEAKNVSTLMKILVQVLSNPK 6221

Search completed: January 3, 2005, 15:25:30
Job time : 131.333 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:53:45 ; Search time 95.3333 Seconds
(without alignments)
1934.128 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 3659

Sequence: 1 MSISDEVNFLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: Genesecp1980s:*
- 2: Genesecp1990s:*
- 3: Genesecp2000s:*
- 4: Genesecp2001s:*
- 5: Genesecp2002s:*
- 6: Genesecp2003as:*
- 7: Genesecp2003bs:*
- 8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3650	99.8	514	8	Adj75513 Marker ge
2	3641	99.5	514	4	Aab95225 Human pro
3	3640	99.5	514	6	Abo07190 Human p53
4	3615	98.8	514	8	Adj76263 Marker ge
5	3355.2	91.7	577	7	Add14051 Human src
6	3355.2	91.7	577	8	Adq18019 Human sof
7	3282.7	89.7	459	4	Admi19860 Protein e
8	3265.6	89.2	542	5	Abp41760 Human ova
9	2817.1	79.7	700	4	Abb60376 Drosophil
10	2816.1	77.0	584	4	Abg21351 Novel hum
11	2744	75.0	395	5	Abp51424 Human MDD
12	2618.9	71.6	5635	5	Abp60991 Novel hum
13	2614.8	71.5	5636	7	Adj70089 Human hea
14	2614.8	71.5	5636	7	Adj83137 Human hem
15	2614.8	71.5	5636	8	Adk60205 Angiogene
16	2614.8	71.5	5636	8	Adk60506 Angiogene
17	2614.8	71.5	5636	8	Adp73129 Angiogene
18	2612	71.4	7064	7	ADG70546 Aspergilli
19	2610.5	71.3	4599	8	Adi27168 Mouse LRP
20	2610.5	71.3	4599	8	Adi27169 Mouse LRP
21	2609.5	71.3	31267	6	Abg74786 Human RGS
22	2593.3	70.9	4796	4	Abbs58665 Drosophil
23	2585.9	70.7	4599	6	Abp56837 Human LRP
24	2585.4	70.7	26926	4	Aau05396 Human tit
25	2585.4	70.7	26926	8	Adq17316 Human sof

ALIGNMENTS

RESULT 1

ADJ75513

ID ADJ75513 standard; protein; 514 AA.

XX AC ADJ75513;

XX DT

XX XX 20-MAY-2004 (first entry)

XX DE

XX XX Marker gene related amino acid sequence SEQ ID NO:765.

XX XX

XX KW bronchial asthma; chronic obstructive pulmonary disease;

XX KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;

XX KW gene therapy; marker.

XX OS Homo sapiens.

XX XX

XX PN EP1394274-A2.

XX XX

XX PD 03-MAR-2004.

XX XX

XX PF 04-AUG-2003; 2003EP-00254857.

XX XX

XX PR 06-AUG-2002; 2002JP-00229312.

XX PR 20-MAR-2003; 2003JP-00077212.

XX XX

XX PA (GENO-) GENOX RES INC.

XX XX

XX PI Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;

XX DR WPI; 2004-193155/19.

XX XX

XX PT Testing for bronchial asthma or chronic obstructive pulmonary disease by

XX PT comparing the expression level of a marker gene in a biological sample

XX PT from a subject with the expression level of the gene in a sample from a

XX PT healthy subject.

XX XX

XX PS Example 11; SEQ ID NO 765; 241pp; English.

XX XX

XX CC The present invention describes a method of testing for bronchial asthma

XX CC or chronic obstructive pulmonary disease. The method comprises

XX CC determining the expression level of a marker gene in a biological sample

XX CC from a subject, comparing the expression level determined with the

XX CC expression level of the marker gene in a biological sample from a healthy

XX CC subject, and judging whether the subject has bronchial asthma or chronic

XX CC obstructive pulmonary disease. The marker gene comprises: (a) a group of

XX CC genes (S1) whose expression levels increase when respiratory epithelial

XX CC cells are stimulated with interleukin-13; or (b) a group of genes (S2)

XX CC whose expression levels decrease when respiratory epithelial cells are

XX CC

XX CC

XX CC

CC stimulated with interleukin-13. Also described: (1) a reagent (I) for
 CC testing for bronchial asthma or chronic obstructive pulmonary disease;
 CC (2) a kit for screening for a candidate compound for a therapeutic agent
 CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
 CC an animal model for bronchial asthma or chronic obstructive pulmonary
 CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
 CC method for producing an animal model for bronchial asthma or chronic
 CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial
 CC asthma or chronic obstructive pulmonary disease, comprising the compound,
 CC a marker gene or an antisense nucleic acid corresponding to a portion of
 CC the marker gene, a ribozyme, a polynucleotide that suppresses the
 CC expression of the gene through an RNAi effect or an antibody recognising
 CC a protein encoded by a marker gene; and (7) a DNA chip for testing for
 CC bronchial asthma or a chronic obstructive pulmonary disease, on which a
 CC probe has been immobilised to assay a marker gene. (I) has respiratory
 CC and antiasthmatic activities, and can be used in gene therapy. The method
 CC is useful for testing for or screening for a therapeutic agent for
 CC bronchial asthma or chronic obstructive pulmonary disease. The present
 CC sequence is used in the exemplification of the present invention.

XX Sequence 514 AA;

Query Match 99.8%; Score 3650; DB 8; Length 514;
 Best Local Similarity 99.6%; Pred. No. 8.4e-56;
 Matches 512; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSISDEVNVLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQGLQYV 60
 DB 1 MSISDEVNVLVRYLQESGFSHSAFTFGIESHSQSNINGALVPPAALISIIQGLQYV 60
 QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 DB 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120
 QY 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFIQAWNPV 180
 DB 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFIQAWNPV 180
 QY 181 SLLASGSGDSTARINWLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 DB 181 SLLASGSGDSTARINWLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
 QY 241 ATGSDGFARITWKDGNLSTLQHGKPIFALKXNKKGNFILSAGVDKTTIWDHTGEA 300
 DB 241 ATGSDGFARITWKDGNLSTLQHGKPIFALKXNKKGNFILSAGVDKTTIWDHTGEA 300
 QY 301 KQFPFHSAAPALVDVQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360
 DB 301 KQFPFHSAAPALVDVQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKISMKQDNCVHDLQOHNKEIYTKWSPGTGNTNPNANLMLASAF 420
 DB 361 GNLLASCSDDMTLKISMKQDNCVHDLQOHNKEIYTKWSPGTGNTNPNANLMLASAF 420
 QY 421 DSTVRLMDVDRGCIHTLTKHQPVSVAFSPPGRVYLASGFKCVHIMNTQTGALVHSY 480
 DB 421 DSTVRLMDVDRGCIHTLTKHQPVSVAFSPPGRVYLASGFKCVHIMNTQTGALVHSY 480
 QY 481 RGTGGIFEVCAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RGTGGIFEVCAAGDKVGASASDGSVCVLDLRK 514

RESULT 2

AAB95225
 ID AAB95225 standard; protein; 514 AA.

XX AAB95225;

XX AAB95225;

DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:17352.

XX

Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the 5602 complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the 5602 complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 514 AA;

Query Match 99.5%; Score 3641; DB 4; Length 514;
 Best Local Similarity 99.4%; Pred. No. 1.2e-55;
 Matches 511; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISDEVNVLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQGLQYV 60

DB 1 MSISDEVNVLVRYLQESGFSHSAFTFGIESHSQSNINGALVPPAALISIIQGLQYV 60

QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120

DB 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQ 120

QY 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFIQAWNPV 180

DB 121 QGSAKNGENTANGEANGAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFIQAWNPV 180

QY 181 SLLASGSGDSTARINWLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240

DB 181 SLLASGSGDSTARINWLSNSTSGTQVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240

Db 181 SLLASGSGDSTARIWNLSENSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
QY 241 ATGSYDGFARIWTKDGNLSTLQHKGPPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300
Db 241 ATGSYDGFARIWTKDGNLSTLQHKGPPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300
QY 301 KQFPFHSAPALDQVMSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
Db 301 KQFPFHSAPALDQVMSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNEIYTIKWSPTGPTGNTNPNANIMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNEIYTIKWSPTGPTGNTNPNANIMLASAF 420
QY 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQTGALVHSY 480
Db 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQTGALVHSY 480
QY 481 RGTGGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514
Db 481 RGTGGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514

RESULT 3
ABO07190
ID ABO07190 standard; protein; 514 AA.
XX
AC ABO07190;
XX
DT 13-AUG-2003 (first entry)
XX
DE Human p53 modifying protein, SEQ ID 150.
XX
KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
KW antapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
KW apoptotic disorder; cell proliferation disorder.
XX
OS Homo sapiens.
XX
PN WO200299122-A1.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002WO-US017382.
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 15-FEB-2002; 2002US-0357253P.
XX
PA (EXEL-) EXELIXIS INC.
XX
PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funk RP;
XX WPI; 2003-156859/15.
DR N-PSDB; ACD13365.
XX
PT Identifying modulators of the p53 pathway for use in treating apoptotic
PT or cell proliferation disorders, comprises screening for agents that
PT modulate activity of a human ortholog of genes that modify the p53
PT pathway in Drosophila.
XX
PS Example 2; Page 458-459; 678pp; English.
XX

The invention relates to identifying (M1) a candidate p53 pathway
modulating agent, by contacting an assay system comprising a purified HM
polypeptide (human ortholog of genes that modify the p53 pathway in
Drosophila) or nucleic acid with a test agent under conditions, where but
for the presence of the test agent, the system provides a reference
activity, and detecting a test agent-biased activity of the assay system.
Also included are modulating (M2) a p53 pathway of a cell (comprising
contacting a cell defective in p53 function with a candidate modulator
that specifically binds to a HM polypeptide comprising an HM amino acid

CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
in a mammalian cell (comprising contacting the cell with an agent that
specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
a disease in a patient (comprising: (a) obtaining a biological sample
from the patient; (b) contacting the sample with a probe for HM
expression; (c) comparing the results with a control; and (d) determining
whether the comparison indicates a likelihood disease). (M1) is useful
for identifying modulators of the p53 pathway. A probe for HM expression
is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
in a patient, where the cancer has greater than 25% expression level.
Modulators identified by (M1) are useful in a variety of diagnostic and
therapeutic applications, where disease or disorder prognosis is related
to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
proliferation disorders (e.g. cancer). Another two new methods (M2 and
M3) are useful for modulating the p53 pathway of a cell, thus restoring
the p53 function of the cell, so that the cell undergoes normal
proliferation or progression through the cell cycle. (M2) and (M3) are
also useful for treating defects in the p53 pathway such as angiogenic,
apoptotic or cell proliferation disorders. The present sequence
represents a human p53 pathway modifying protein

XX
SQ Sequence 514 AA;

Query Match 99.5%; Score 3640; DB 6; Length 514;
Best Local Similarity 99.4%; Pred. NO. 1.3e-55;
Matches 511; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISDEVNFLVRYLOESGFSHAFPGIKSHISOSNINGALVPPAALISIIKGLQYV 60
Db 1 MSISDEVNFLVRYLOESGFSHAFPGIKSHISOSNINGALVPPAALISIIKGLQYV 60
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAAASQ 120
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAAASQ 120
QY 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNAKAVLRGHESEVEFICAWN 180
Db 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNAKAVLRGHESEVEFICAWN 180
QY 181 SLLASGSGDSTARIWNLSENSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
Db 181 SLLASGSGDSTARIWNLSENSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
QY 241 ATGSYDGFARIWTKDGNLSTLQHKGPPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300
Db 241 ATGSYDGFARIWTKDGNLSTLQHKGPPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300
QY 301 KQFPFHSAPALDQVMSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
Db 301 KQFPFHSAPALDQVMSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNEIYTIKWSPTGPTGNTNPNANIMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNEIYTIKWSPTGPTGNTNPNANIMLASAF 420
QY 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQTGALVHSY 480
Db 421 DSTVRLWDVDRGICHTLTKHQPVPVSVAFSPDGRYLASGSPDKCVHIWNTQTGALVHSY 480
QY 481 RGTGGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514
Db 481 RGTGGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514

RESULT 4
ADJ76263
ID ADJ76263 standard; protein; 514 AA.
XX
AC ADJ76263;
XX
DT 20-MAY-2004 (first entry)
XX
DE Marker gene related amino acid sequence SEQ ID NO:1515.

XX bronchial asthma; chronic obstructive pulmonary disease;
 KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;
 KW gene therapy; marker.
 XX Mus musculus.
 XX EF1394274-A2.
 XX 03-MAR-2004.
 XX 04-AUG-2003; 2003EP-00254857.
 XX 06-AUG-2002; 2002JP-00229312.
 XX 20-MAR-2003; 2003JP-00077212.
 XX (GENO-) GENOX RES INC.
 XX Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;
 XX WPI; 2004-193155/19.
 XX Testing for bronchial asthma or chronic obstructive pulmonary disease by
 PT comparing the expression level of a marker gene in a biological sample
 PT from a subject with the expression level of the gene in a sample from a
 PT healthy subject.
 XX Claim 16; SEQ ID NO 1515; 241pp; English.
 XX The present invention describes a method of testing for bronchial asthma
 CC or chronic obstructive pulmonary disease. The method comprises
 CC determining the expression level of a marker gene in a biological sample
 CC from a subject, comparing the expression level determined with the
 CC expression level of the marker gene in a biological sample from a healthy
 CC subject, and judging whether the subject has bronchial asthma or chronic
 CC obstructive pulmonary disease. The marker gene comprises: (a) a group of
 CC genes (S1) whose expression levels increase when respiratory epithelial
 CC cells are stimulated with interleukin-13; or (b) a group of genes (S2)
 CC whose expression levels decrease when respiratory epithelial cells are
 CC stimulated with interleukin-13. Also described: (1) a reagent (I) for
 CC testing for bronchial asthma or chronic obstructive pulmonary disease;
 CC (2) a kit for screening for a candidate compound for a therapeutic agent
 CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
 CC an animal model for bronchial asthma or chronic obstructive pulmonary
 CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
 CC method for producing an animal model for bronchial asthma or chronic
 CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial
 CC asthma or chronic obstructive pulmonary disease, comprising the compound,
 CC a marker gene or an antisense nucleic acid corresponding to a portion of
 CC the marker gene, a ribozyme, a polynucleotide that suppresses the
 CC expression of the gene through an RNAi effect or an antibody recognising
 CC a protein encoded by a marker gene; and (7) a DNA chip for testing for
 CC bronchial asthma or a chronic obstructive pulmonary disease, on which a
 CC bronchial asthma or a chronic obstructive pulmonary disease, (I) has respiratory
 CC and antiasthmatic activities, and can be used in gene therapy. The method
 CC is useful for testing for or screening for a therapeutic agent for
 CC bronchial asthma or chronic obstructive pulmonary disease. The present
 CC sequence is used in the exemplification of the present invention.
 XX Sequence 514 AA;

Query Match 98.8%; Score 3615; DB 8; Length 514;
 Best Local Similarity 98.4%; Pred. No. 3.9e-55;
 Matches 506; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSISDSEWFLVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISITKGLQYV 60
 DB 1 MSISDSEWFLVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISITKGLQYV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAAASQ 120
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAAATNQ 120

QY 121 QGSAKNGENTANGEAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWN 180
 DB 121 QGSAKNGENTANGEAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFICAWN 180
 QY 181 SLLASGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 DB 181 SLLVSGSGDSTARINWLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 QY 241 ATGSDGPFARIWTKOGLASTLGOHKGPILFALKWKNKGNFILLSAGVDKTTIWDHAHTGEA 300
 DB 241 ATGSDGPFARIWTKOGLASTLGOHKGPILFALKWKNKGNFILLSAGVDKTTIWDHAHTGEA 300
 QY 301 KOQFPFHSAPALVDVWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 DB 301 KOQFPFHSAPALVDVWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWMSKODNCVHDLQOHNKEIYTIKWSPTGPTGNTNPNANLMLASAF 420
 DB 361 GNLLASCSDDMTLKIWMSKODNCVHDLQOHNKEIYTIKWSPTGPTGNTNPNANLMLASAF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLASGDFKCVHIWNTQTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLASGDFKCVHIWNTQTGALVHSY 480
 QY 481 RTGGGIFECVWNAAGDKVGASASDGSVCVLDLRK 514
 DB 481 RTGGGIFECVWNAAGDKVGASASDGSVCVLDLRK 514

RESULT 5

ADD14051
 ID ADD14051 standard; protein; 577 AA.

AC ADD14051;

DT 01-JAN-2004 (first entry)

DE Human src biomarker polypeptide SEQ ID NO:240.

XX predictor set; protein tyrosine kinase activity modulator;

KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;

KW gene therapy; drug sensitivity; genetic profile; cancer; human.

XX Homo sapiens.

XX WO2003062395-A2.

XX 31-JUL-2003.

PF 17-JAN-2003; 2003WO-US001981.

XX 18-JAN-2002; 2002US-0350061P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Huang F, Fairchild CR, Lee FY, Shaw P;

XX WPI; 2003-636735/60.

DR N-PSDB; ADD14646.

XX New polynucleotides and polypeptides for predicting the activity of

PT compounds that interact with protein tyrosine kinases and/or protein

PT tyrosine kinase pathways.

XX Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and

correlating the expression of the markers to the compound's ability to modulate the activity of the cells; (2) a plurality of cell lines for identifying polynucleotides and polypeptides whose expression levels correlate with compound sensitivity or resistance of cells associated with a disease state; and (3) identifying polynucleotides and polypeptides that predict compound sensitivity or resistance of cells associated with a disease state; and (3) identifying polynucleotides and polypeptides that predict compound sensitivity or resistance of cells associated with a disease state by using the expression pattern of the microarray. The polynucleotides and polypeptides have cytoskeletal activities, and can be used in gene therapy. The polynucleotides and polypeptides are useful in predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways. These may be used in determining drug sensitivity in patients to allow the development of individualized genetic profiles which aid in treating diseases and disorders (e.g. cancer) based on patient response at a molecular level. The present sequence is used in the exemplification of the present invention.

XX Sequence 577 AA;

Query Match 91.7%; Score 3355.2; DB 7; Length 577;
 Best Local Similarity 86.2%; Pred. No. 6.2e-50;
 Matches 456; Conservative 38; Mismatches 17; Indels 18; Gaps 10;

QY 1 MSISDEVNPLVRYLQESGFSHAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
 DB 52 MSITSDEVNPLVRYLQESGFSHAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 111
 QY 61 EAEVSINEDGTLPDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQQ--AAAAAAA-AA 117
 DB 112 EAEVSINEDGTLPDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQQ--AAAAAAA-AA 171
 QY 118 ASO-----QG-S-----AKNGENTANGAHTIANNHTDM-MEVDGDVEIPPNKAVLVR 166
 DB 172 ATAATTTSAGVSHQNPKNREATVNGEENRAHSV-NNHA-KPMEIDGEVEIPSKATVLR 229
 QY 167 GHESEVFIICAWNPVSDLLASGSDSTARIWNLSENSTGSLVLRHCIREGGQDVPSNK 226
 DB 230 GHESEVFIICAWNPVSDLLASGSDSTARIWNLSENSTGSLVLRHCIREGGQDVPSNK 289
 QY 227 DVTSLDWNSBGTLLATGSDYDGFARIVTKGNLSTLQHKGPITFALKKWKNGFIISAGV 286
 DB 290 DVTSLDWNTGTLATGSDYDGFARIVTKGNLSTLQHKGPITFALKKWKNGFIISAGV 349
 QY 287 DKTTIIWDAHTGEAKQOFPFHSAPALDQVQSN--TPFASCTDMCHVCKLGDPRPKTF 345
 DB 350 DKTTIIWDAHTGEAKQOFPFHSAPALDQVQ--NNTTFASCTDMCHVCKLGDPRPKTF 408
 QY 346 QGHTNEVNAIKWDPPTGNLLASCDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPG 405
 DB 409 QGHTNEVNAIKWDPPTGNLLASCDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468
 QY 406 TNNPNANMLASFSFSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKC 465
 DB 469 TSNPNINMLASFSFSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSFDKC 528
 QY 466 VHIWNTQTGLVHYSYRGTTGGIPVPCWNAAGDKVGSASDGSVCVLDLRK 514
 DB 529 VHIWNTQSGNLVHYSYRGTTGGIPVPCWNAAGDKVGSASDGSVCVLDLRK 577

RESULT 6
 ADQ18019
 ID ADQ18019 standard; protein; 577 AA.

XX ADQ18019;

AC ADQ18019;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated protein - SEQ ID 836.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
 XX Homo sapiens.
 XX W02004048938-A2.
 XX 10-JUN-2004.
 XX 26-NOV-2003; 2003WO-US038193.
 XX 26-NOV-2002; 2002US-0429739P.
 XX (PROT-) PROTEIN DESIGN LABS INC.
 XX Aziz N, Ginsburg WM, Zlotnik A;
 XX WPI; 2004-441208/41.
 XX Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.
 XX Example 2; SEQ ID NO 836; 210pp; English.

The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytoskeletal applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated protein of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

Sequence 577 AA;

Query Match 91.7%; Score 3355.2; DB 8; Length 577;
 Best Local Similarity 86.2%; Pred. No. 6.2e-50;
 Matches 456; Conservative 38; Mismatches 17; Indels 18; Gaps 10;

QY 1 MSISDEVNPLVRYLQESGFSHAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
 DB 52 MSITSDEVNPLVRYLQESGFSHAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 111
 QY 61 EAEVSINEDGTLPDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQQ--AAAAAAA-AA 117
 DB 112 EAEVSINEDGTLPDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQQ--AAAAAAA-AA 171
 QY 118 ASO-----QG-S-----AKNGENTANGAHTIANNHTDM-MEVDGDVEIPPNKAVLVR 166
 DB 172 ATAATTTSAGVSHQNPKNREATVNGEENRAHSV-NNHA-KPMEIDGEVEIPSKATVLR 229
 QY 167 GHESEVFIICAWNPVSDLLASGSDSTARIWNLSENSTGSLVLRHCIREGGQDVPSNK 226
 DB 230 GHESEVFIICAWNPVSDLLASGSDSTARIWNLSENSTGSLVLRHCIREGGQDVPSNK 289
 QY 227 DVTSLDWNSBGTLLATGSDYDGFARIVTKGNLSTLQHKGPITFALKKWKNGFIISAGV 286
 DB 290 DVTSLDWNTGTLATGSDYDGFARIVTKGNLSTLQHKGPITFALKKWKNGFIISAGV 349
 QY 287 DKTTIIWDAHTGEAKQOFPFHSAPALDQVQSN--TPFASCTDMCHVCKLGDPRPKTF 345
 DB 350 DKTTIIWDAHTGEAKQOFPFHSAPALDQVQ--NNTTFASCTDMCHVCKLGDPRPKTF 408
 QY 346 QGHTNEVNAIKWDPPTGNLLASCDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPG 405
 DB 409 QGHTNEVNAIKWDPPTGNLLASCDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468

PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476159/51.
XX N-PSDB; ADML9381.
XX
PT Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX Claim 11; SEQ ID NO 667; 809pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule encoding a
CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to a protein of the
CC invention.
XX
XX Sequence 459 AA;
SQ
Query Match 89.7%; Score 3282.7; DB 4; Length 459;
Best Local Similarity 92.9%; Pred. No. 5e-49;
Matches 457; Conservative 2; Mismatches 0; Indels 33; Gaps 3;
QY 23 HSAFTFGIKSHISQSNINGALVPPAALISIQKLOQVVEAEVSNEDGTLFDCGRPIESLS 82
Db 1 H-A-----SD-----VEAEVSNEDGTLFDCGRPIESLS 27

QY 83 LIDAVNPVVQTRQAYRDKLAQQQAAAAAAAAAAAAAQQGSAKNGENTANGENGATIA 142
Db 28 LIDAVNPVVQTRQAYRDKLAQQQAAAAAAAAAAAAAQQGSAKNGENTANGENGATIA 87
QY 143 NNHTDMMEVDGVEIIPPKNKAVLVRGHESEVFIICAMPVSDLLASGSGDSTARINWLSNS 202
Db 88 NNHTDMMEVDGVEIIPPKNKAVLVRGHESEVFIICAMPVSDLLASGSGDSTARINWLSNS 147
QY 203 TSGSTOLVLRHCIREGGQDVPSNKNVTSIDMNSSEGTLLATGSDYDGFARIWTKDGNLSTL 262
Db 148 TSGSTOLVLRHCIREGGQDVPSNKNVTSIDMNSSEGTLLATGSDYDGFARIWTKDGNLSTL 207
QY 263 GQHKGFIFALKWNNKGNFILSAGVDKTTIWDHTGEAKQOQFFHSPALDLDVQSNNTF 322
Db 208 GQHKGFIFALKWNNKGNFILSAGVDKTTIWDHTGEAKQOQFFHSPALDLDVQSNNTF 267
QY 323 ASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGTGNLLASCSDMTLKIWSMKQDN 382
Db 268 ASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGTGNLLASCSDMTLKIWSMKQDN 327
QY 383 CVHDLQHNKEIYTIKWSPTGCTNNPNANMLASASFSSTVRLWDVDRGICHTLTQH 442
Db 328 CVHDLQHNKEIYTIKWSPTGCTNNPNANMLASASFSSTVRLWDVDRGICHTLTQH 387
QY 443 EPVYSVAFSPDGRYLASGSGFDKCVHIWNTQTGALVHSYRGTGGIFEVYVNAAGDKVGASA 502
Db 388 EPVYSVAFSPDGRYLASGSGFDKCVHIWNTQTGALVHSYRGTGGIFEVYVNAAGDKVGASA 447
QY 503 SDGSVCVLDLRK 514
Db 448 SDGSVCVLDLRK 459
RESULT 8
ABP41760
ID ABP41760 standard; protein; 542 AA.
XX
AC ABP41760;
XX
DT 22-AUG-2002 (first entry)
XX Human ovarian antigen HNOKM38, SEQ ID NO:2892.
XX Human; ovarian antigen; ovary; ovary; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive.
XX Homo sapiens.
XX
XX WO200200677-A1.
XX
XX 03-JAN-2002.
XX
XX 07-JUN-2001; 2001WO-US018569.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
XX WPI; 2002-147878/19.
XX N-PSDB; ABQ54837.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT

Db 115 NQNNNAKPEIKBPTGTGAGAGCNKIAGSTTCTSTPTDQSAASEVDSNGNAANAGGTVA 174
QY 113 ----A-----A-----AA-ASO-----Q-----GSAK--N-G-EN--T-- 130
Db 175 GNNAGAGNOASTGCGNSTSTPAGDGLAAPGASOKKQNSNEAGSSGSGNAGNANATSTDD 234
QY 131 -A-----NG-----EE-----N-----G-AH-----TI- 141
Db 235 AASSTSTNGNSSSSVEQPTSLGTAGGTVSTSNPDAAASGASGATGSKAPSGAVTIR 294
QY 142 --A--NN-----H-----T-DMEV--DGD--VRIPPKXAVVLRG 167
Db 295 VQAQGNVQSGSSNAOSSAPSGTISSTSGGAGTAAAL-VPMDIDENIPIESKARVLRG 353
QY 168 HSEVEFICAWNPVS-DLLASGSDSTARIWNLSE-NSTSGSTOLVLRHCIREGQDVPSN 225
Db 354 HSEVEFICAWNP-SRDLASGSDSTARIWNLSE-NSTSGSTOLVLRHCIREGQDVPSN 410
QY 226 KDVTSLDMNSEGLLATGSDGPARIW-TKDGNLASTLGHKGPFPALKWKK-GNFILS 283
Db 411 KDVTSLDMNDCGSLLATGSDGPARIWKT-DGRLASTLGHKGPFPALKWKN-KCGNYLLS 468
QY 284 AGVDKTTIWDATGE-AKQOPFPHPAPALDVQSNNTFASGSTD-MCHVCKLG--QD 339
Db 469 AGVDKTTIWDASTGQCT-QQAFHAPALDVQSNNTFASGSTDQD-IHVCRLGWNE- 525
QY 340 RPIKTFQHTNEVNAIKWDPNTGNLLASCSDDMTLKIWSMKOD-NCVHDLQOHKKEIYTIK 398
Db 526 -PIKTFQHTNEVNAIKWCPQGLLASCSDDMTLKIWSMRDRC-HDLQAHKKEIYTIK 583
QY 399 WSPGTGTTNNPNANMLASASFDSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLA 458
Db 584 WSPGTGTTNNPNANMLASASFDSTVRLWVDVDRGICHTLTKHQEPVYSVAFSPDGRYLA 643
QY 459 SSGFDCVCHVWNTQTGALVHSYRGTCGIFEVCMNAAGDKVGSASDGSVCVLDLRK 514
Db 644 SSGFDCVCHVWNTQTGALVHSYRGTCGIFEVCMNAAGDKVGSASDGSVCVLDLRK 699

RESULT 10
ABG21351
ID ABG21351 standard; protein; 584 AA.
XX AC ABG21351;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #21342.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US0008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSB-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PI WPI: 2001-639362/73.
XX DR N-PSDB; AAS85538.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations.

PT responsible for genetic disorders or other traits and to assess
XX biodiversity.
PS Claim 20; SEQ ID NO 51710; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AEG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 584 AA;
Query Match 77.0%; Score 2816.1; DB 4; Length 584;
Best Local Similarity 63.8%; Pred. No. 1.3e-39;
Matches 387; Conservative 44; Mismatches 37; Indels 139; Gaps 54;
QY 2 SI-----S-----S-----D-----EV---N---FLV-YRY-LOE-SGFSSHATFGIKSH 33
Db 22 AIGAVGRSOASTLAGQPDGNGVCVHAGNSRIF-VPLPVGL-EVKGFSHATFGIESH 79
QY 34 ISQSNINGALVPPAALISIIQGLQYVEAEVSNEDGTLFDGRPIESLSLIDAVMPDVQ 93
Db 80 ISQSNINGTLVPPSALISILQGLQYVEAEISINKDGTVDSPRIESLSLIVAVPDVQ 139
QY 94 TRQAYRDKLAQQA-AAA--A-A-AAA-----AA-SQGSANNGENTANGENGATHTIA 142
Db 140 MRQQAQGEKLTQQAASAAATEASAKAATMTAPALISQNPKNREATVNGEANGAEI- 198
QY 143 NNHTDM-MEVDGDVEIPPNKAVLRGHESEVFCANPNVSDLLASGSGSTARIWNLSEN 201
Db 199 NNHS-KPMEIDGDVEIPPNKATVLRGHESEVFCANPNVSDLLASGSGSTARIWNLSEN 257
QY 202 STSGSTQLVLRHCIREGQDVPSNKDVTSLDWNSEGTLLATGSDYDGFARITWKDGNLAST 261
Db 258 SNGSGTQLVLRHCIREGHDVPSNKDVTSLDWNSDGTLLAMGSDYDGFARITWTE--N-A-- 312
QY 262 LGOHKGPFPAL--KW-NKKGN--FILS-AGVDKTTIWDATHT---GEAKQOPFPHSAPAL 312
Db 313 -----P--ALDVDMQN---NMTFA-SCS-TDM--CI---HVCRLGCD-----H--P-- 344
QY 313 DVD-WQSNNTFASCSDDMTLKIWSMKOD-NCVHDLQOHKKEIYTIK-----QG- 347
Db 345 -VKTFQ-GHTF--C-T--CIESICFWGGLRKLKTTMTTEGKRLRP-KTFCSDGGALLPPAGR 396
QY 348 --H--T-----NEVNAIKWDPNTGNLLASCSDDMTLKIWSMKODNCVHDLQO 389
Db 397 RPHLLTGPDIPFKRLSALFQNEVNAIKWDPNSGMLLASCSDDMTLKIWSMKODACVHDLQA 456
QY 390 HNKKEIYTIKWSPTGPTNNPNANMLASASFDSTVRLWVDVDRGICHTLTKHQEPVYSVA 449
Db 457 HSKKEIYTIKWSPTGPTNNPNANMLASASFDSTVRLWVDVDRGICHTLTKHQEPVYSVA 516
QY 450 FSPDGRYLAGSGFDCVCHVWNTQTGALVHSYRGTCGIFEVCMNAAGDKVGSASDGSVCV 509
Db 517 FSPDGRYLAGSGFDCVCHVWNTQTGALVHSYRGTCGIFEVCMNAAGDKVGSASDGSVCV 576

Qy 510 L-DLR-K 514
 Db 577 LWQLSGK 583

RESULT 11
 ABP51424
 ID ABP51424 standard; protein; 395 AA.
 XX AC ABP51424;
 XX DT 03-SEP-2002 (first entry)
 XX DE Human MDDT SEQ ID NO 446.
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
 KW antiallergic; antianaemic; antiaethmatic; antiatherosclerotic; antiout;
 KW neuroprotective; antirheumatic; antiarthritic.
 XX OS Homo sapiens.
 XX PN WO200240715-A2.
 XX PD 23-MAY-2002.
 XX PF 06-SEP-2001; 2001WO-US027628.
 XX PR 05-SEP-2000; 2000US-0229747P.
 XX PR 05-SEP-2000; 2000US-0229748P.
 XX PR 05-SEP-2000; 2000US-0229749P.
 XX PR 05-SEP-2000; 2000US-0229750P.
 XX PR 05-SEP-2000; 2000US-0229751P.
 XX PR 06-SEP-2000; 2000US-0230583P.
 XX PR 06-SEP-2000; 2000US-0230505P.
 XX PR 06-SEP-2000; 2000US-0230514P.
 XX PR 06-SEP-2000; 2000US-0230515P.
 XX PR 06-SEP-2000; 2000US-0230518P.
 XX PR 06-SEP-2000; 2000US-0230519P.
 XX PR 06-SEP-2000; 2000US-0230595P.
 XX PR 06-SEP-2000; 2000US-0230597P.
 XX PR 06-SEP-2000; 2000US-0230598P.
 XX PR 06-SEP-2000; 2000US-0230599P.
 XX PR 06-SEP-2000; 2000US-0230610P.
 XX PR 06-SEP-2000; 2000US-0230615P.
 XX PR 06-SEP-2000; 2000US-0230988P.
 XX PR 06-SEP-2000; 2000US-0230989P.
 XX PR 07-SEP-2000; 2000US-0230951P.
 XX PR 07-SEP-2000; 2000US-0231163P.
 XX PR 07-SEP-2000; 2000US-0231167P.
 XX PA (INCY-) INCYTE GENOMICS INC.
 XX PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;
 PI Hallman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;
 PI Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;
 PI Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;
 PI Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
 DR WPI; 2002-527544/56.
 DR N-PSDB; ABQ72641.
 XX Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
 PT AIDS.
 XX PS Claim 14; Page 578; 618pp; English.
 XX

CC The invention relates to an isolated human disease detection and
 CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I). (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional MDDT. (I)
 CC or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDDT, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germline gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences
 XX SQ Sequence 395 AA;

Query Match 75.0%; Score 2744; DB 5; Length 395;
 Best Local Similarity 99.5%; Pred. No. 4.7e-39;
 Matches 389; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSISDEVNPLVYRYLOESGFSHSAFTGKSHISQSNINGALVPPAALISITQKLOYV 60
 Db 5 MSISDEVNPLVYRYLOESGFSHSAFTGKSHISQSNINGALVPPAALISITQKLOYV 64
 Qy 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVOTROQAVRDKLAQOAAAAAASQ 120
 Db 65 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVVOTROQAVRDKLAQOAAAAAASQ 124
 Qy 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFCANPV 180
 Db 125 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFCANPV 184
 Qy 181 SLLASGSGDSTARINWLSNSTSGSTQVLRHCIREGGQDVPSNKDVTSLDWSNSETLL 240
 Db 185 SLLASGSGDSTARINWLSNSTSGSTQVLRHCIREGGQDVPSNKDVTSLDWSNSETLL 244
 Qy 241 ATGSDGFARITWKDGNLSTLQHKGPFPALKWKNKGNFILSAGVDKTTIWDATGEA 300
 Db 245 ATGSDGFARITWKDGNLSTLQHKGPFPALKWKNKGNFILSAGVDKTTIWDATGEA 304
 Qy 301 KQQFPFHSAPALVDVWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 Db 305 KQQFPFHSAPALVDVWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 364
 Qy 361 GNLLASCSDDMTLKIKWSMKQDNCVHDLQOHN 391
 Db 365 GNLLASCSDDMTLKIKWSMKQDNCVHDLQOHN 395

RESULT 12
 ABP60991
 ID ABP60991 standard; protein; 5635 AA.
 XX AC ABP60991;
 XX

DT 10-SEP-2002 (first entry)
DE Novel human protein. SEQ ID 78.
XX
XX Human; cytostatic; vulnary; antiarteriosclerotic; antiparkinsonian;
KW neurotropic; neuroprotective; immunosuppressive; haemostatic;
KW antinflammatory; cardiant; antiulcer; virucide; antithyroid;
KW cerebroprotective; anorectic; metabolic; vaccine; cancer; infection;
KW wound healing disorders; atherosclerosis; Parkinson's disease;
KW Alzheimer's disease; autoimmune disorder; haematopoietic disorder;
KW inflammation; neoplastic disease; nervous system disorder;
KW cardiovascular disorders; pancreatitis; respiratory disorder;
KW hyperproliferation; systemic autoimmune disease; hyper-immunity;
KW developmental abnormality; gastrointestinal ulceration; neuropathy;
KW haematological disease; metabolic disease; sperm dysfunction;
KW thyroid disorder; hypothyroidism; brain damage; colitis;
KW cone photo-transduction deficiency; neurological disease; stroke;
KW angogenesis; ovulation disorder; spinal cord; thyroid gland; heart;
KW trachea; thymus; lymph node; muscular system; obesity; anorexia;
KW growth abnormality; precocious puberty.
XX
OS Homo sapiens.
XX
XX WO200250105-A1.
XX
XX 27-JUN-2002.
XX
XX 17-DEC-2001; 2001WO-US049232.
XX
XX 19-DEC-2000; 2000US-0256710P.
XX 20-DEC-2000; 2000US-0257048P.
XX 09-JAN-2001; 2001US-0260482P.
XX 30-JAN-2001; 2001US-0264922P.
XX 06-FEB-2001; 2001US-0266797P.
XX 19-MAR-2001; 2001US-0276988P.
XX 04-APR-2001; 2001US-0281535P.
XX 08-MAY-2001; 2001US-0289622P.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
XX (GLAX) GLAXO GROUP LTD.
XX
XX Agarwal P, Birkeland M, Cogswell JP, Kabnick KF, Lai Y;
XX Martensen SA, Rizvi SK, Smith RF, Strum JC, Xie Q;
XX
XX WPI: 2002-508784/54.
XX N-PSDB; ABQ86156.
XX
XX Secreted proteins and polynucleotides useful as vaccines for preventing
XX or treating various diseases e.g. cancer, wounds, atherosclerosis,
XX Parkinson's disease, Alzheimer's disease, infection, autoimmune disorder.
XX
XX Claim 1(a); Page 285-297; 335pp; English.
XX
XX The invention relates to an isolated polypeptide with signal sequences
XX which allow it to be secreted extracellularly or membrane associated. The
XX activity of polypeptides of the invention may be described as,
XX cytostatic, vulnary, antiarteriosclerotic, antiparkinsonian, neurotropic,
XX neuroprotective, immunosuppressive, haemostatic, antinflammatory,
XX cardiant, antiulcer, virucide, antithyroid, cerebroprotective, anorectic,
XX and metabolic. Polypeptides and polynucleotides of the invention are
XX useful in the treatment, or as a vaccine in the prevention of, cancer,
XX wound healing disorders, infection, atherosclerosis, Parkinson's disease
XX and Alzheimer's disease, autoimmune disorder, haematopoietic disorder,
XX inflammation, neoplastic diseases, nervous system related disorders and
XX cardiovascular disorders, pancreatitis, respiratory disorder,
XX hyperproliferation, systemic autoimmune disease, hyper-immunity,
XX developmental abnormality, gastrointestinal ulceration, neuropathy,
XX haematological diseases, metabolic diseases, sperm dysfunction, thyroid
XX disorders e.g. hypothyroidism, brain damages, colitis, cone photo-
XX transduction deficiency, neurological diseases, stroke, angiogenesis,
XX ovulation disorders, diseases in the spinal cord, thyroid gland, heart,
XX trachea, thymus, lymph node and muscular system, obesity, anorexia,

CC growth abnormalities, and alleviation of precocious puberty. The
CC sequences given in records ABP60965-ABP61019 represent novel human
CC proteins of the invention
XX
XX Sequence 5635 AA;

Query Match 71.6%; Score 2618.9; DB 5; Length 5635;
Best Local Similarity 9.6%; Pred. No. 4.4e-31;
Matches 440; Conservative 59; Mismatches 13; Indels 4071; Gaps 372;

QY 1 M-S-----I-----SS--D-----EV-----8
DB 1 MISWEVHTVFLFALLYSLAQDASPSQSEIRAEIPEGASTLAFVDPDTGSMYDVLQVI 60
QY 9 -----NF-LV-----12
DB 61 EGASKILETSKRPKPLFNFALVFPFHDPEIGPVTITDPKFKFQYELRELYVQGGDCPE 120
QY 13 -----Y-----R-----Y-----LQ-----17
DB 121 MSIGAIAKIALEISLPGSFYVFTDARSKDYRLTHEVLQIQKQSQVFLTGDCDDRTH 180
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
DB 181 IGYKVEIASTSGQVFLHDKKQVNEVLKQVEEAVQASKVHLLSTDHLEQAVNTWRIFP 240
QY 27 -----T-----FG-----I-----K 31
DB 241 DFLKEVTVLSGSPFMEIRNPLGKLIKKGGLHLLHNSAKVVNVKEPEAGNWTVK 300
QY 32 ---S--H-----I-----S--Q-----36
DB 301 TSSSGRHRSVRITGLSTIDFRAGFSRKPTLDFKKTVSFVQGIPTYVLLNTSGISTPARID 360
QY 37 -----S-----NI-----N-----G-----41
DB 361 LLELLISGSSUKTIPVKYYPHRKPYGIWNISDFVPPNEAFFLKVGTGDKDDYLFORVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
DB 421 VSFSSIVDPAPKVTMPKPTPGYVLPQGPICSDVSLPFTLSFVRNGVTLGVDDYQLKESA 480
QY 48 -----A--L-----I-----S-----IIQ-----KG-----56
DB 481 SVNLDIAKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPPIQVNNVTVPGERAVLT 540
QY 57 -----L--Q-----Y-----V--E-----61
DB 541 CLIIISAVDYNLTQWRNDRDLRLAEPARINTLANLSLELKVFNDAAGEYHCVSSEGGSS 600
QY 62 -A-----EVS1-----NE-----68
DB 601 AASVFLTVQEPKVTMPKQSFQSGSEVSIMCSATGYPKPKIAVTNDFIVGSHRYRM 660
QY 69 --DGLTF-----D-----G-----R-----P-----I--78
DB 661 TSDGTLFIKNAAPKAGIYVGLASAGTDKQNSTLRYIEAPKLMVQSELLVALGDITV 720
QY 79 -E--S-----L--S--L-----DA-----V-----87
DB 721 MECKTSGIPPPQVQKFWKPGDLRLPSTFLIDPLLLGLLKIQETQDLQDAGDYTCVAINAGR 780
QY 88 -----M-----P-DV-----VQ-----T-----94
DB 781 ATGKITLDVGSPPVFIQEPADVMEIGSNVTLPCYVQGPETIKWRRLDNWPIFSRPPS 840
QY 95 -----R-----95
DB 841 VSSISQLRTGALFILNLWASDKGTICEAENQFGKIQSETTIVTGTGLVAPLIGISPSVAN 900
QY 96 ---QQ-----A-----Y-----R-----D--KL--103
DB 901 VIEGQQLTLPCTLLAGNPFPERRWIKNSAMLLQNPYITVRSGLSILIERVQLQDQGEYTC 960

QY 104 A Q-QO AA 109
Db 961 VASVAGTNKTSVVHVLPITQHQOILSTIEGIPVTLPCASGNPKPSVIMSKGEL 1020
QY 110 AAA A A 118
Db 1021 ISTSAKFSAGADSLVYVSPGSEGEYVCTATNTAGYAKRKVOLTYYVPRVFGDORG 1080
QY 119 SQ Q GS 123
Db 1081 LSQDKPVEISVLAGEEVLPCFVKSLPPIITWAKETQLISPFSPRHTFLPSCMKITET 1140
QY 124 A K N 126
Db 1141 RTSDSGMYLCVATNIAGNVQAVKLVNVHPKIQRGPKHLKVQGVQGVDPICNAQGTPLP 1200
QY 127 GE N T A N G 133
Db 1201 VITWSKGGTMLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDEITLHVQE 1260
QY 134 E B N GA H 139
Db 1261 PPTVEDLEPPYNTTFOERVANQRIEFPCKPAPKTPKTIKWLHNGRELTGREPGISILEDG 1320
QY 140 T IA N N HT DM 148
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHPVVIKQEQVNVSVLLNQLTNL 1380
QY 149 M N E 150
Db 1381 FCEVEGTPSPIMMYKDNQVQTESSTIQTVNNGKILKLPRATPEDAGRYSCAKAINAGTS 1440
QY 151 VD G DV B I P PN 160
Db 1441 QKYFNIDVLVPTTIIGTFNFEVSVLNRDVALEQVKGTFPDIHWFKDGKFLGDPN 1500
QY 161 K A V VL 165
Db 1501 VELLDRGOVLHLKNARNDKRGVQCTVSNAGKQAKDKIKTIYNPPSIKGNVTTDISVL 1560
QY 166 RG H 168
Db 1561 INSLIKLECETRGLPMPAITWYKDGQPIMSQALYIDKQYLHIAPRAQVSDSATYCHV 1620
QY 169 E S VFI C A W 177
Db 1621 ANVAGTAKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECKAAGNPSPILTWLKGV 1680
QY 178 N P V SD 182
Db 1681 PVKANDNIRIEAGGKLEINSAQEI DRGOYICVATSVAGEKEIKYEVDVLVPAIEGGDE 1740
QY 183 LL ASGS G D S 191
Db 1741 TSYFVIMVNNLLDCHVTGSPPTIMWLKQGLIDBERDGFILLNGRKLVIQAQVSNVT 1800
QY 192 TA R I W 196
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSSGLSERVVVKYKPVALQCIANGIPNSITW 1860
QY 197 NL S 199
Db 1861 LKDDQPNTAQNKLKIOSSGRVLQIAKLLLEDAGRYTCVATNAAGETOQHILHVHEPPS 1920
QY 200 E N S T 207
Db 1921 LEDAGKMLNLTVLVSNFVQLEKAGNPVPVITWYKDNRLLSGTSMTFLNRQIIDIES 1980
QY 208 Q L V LR H C I 215
Db 1981 AQISDAGIYKCVAINSAGATELFSLQVHVAPSI SGSNMNVAVVNNPVLECEARGIPA 2040

QY 216 R E G Q O D 221
Db 2041 PSUTWLKOGSPVSSFNGQLVLSGGRIALATSAQISDTRYTCVAVNAAGEKORDIDLRV 2100
QY 222 VPSN KD VT S L 231
Db 2101 YVPPNMGBEQNVSVLISQAVELLQSDAIPPTLTWLKDGHPLLKPKGLSISENSRLVK 2160
QY 232 D W N S EG TLL A 241
Db 2161 IEDAQVQDTRYTCVATNAGTKERNYVNIWPPNIGSGDELQTLTVIEGNLISLICES 2220
QY 242 T GS Y D G F A 249
Db 2221 SGIPPPNLTKKKGSPVLTDMSGRVRLISGGROLQISIAEKSDAALYSCVASNVAGTAKK 2280
QY 250 R I WTGD 256
Db 2281 EYNLQVIRPTITNSGSHPEIIVTRGKSI SLECEVOGIPPTVTWTKDGHPLIKAKGVE 2340
QY 257 NL AS T L GQH 265
Db 2341 ILDEGHILQKNIHVSQDTRYCVAVNAGMTDKKYDLSVHAPPSIIGNHRSPENISVVE 2400
QY 266 K G P I FALK W 274
Db 2401 KNSVSLTCEASGIPLSITWF KOGWPVLSNSVRLISGGRMLRLMQTTMEDAGQYTCV 2458
QY 275 N K K GN 279
Db 2459 VRNAAGEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVEITWHDQ 2518
QY 280 F I L S AGV 286
Db 2519 PLOEDBAHHIISGGRFLQITNVQVPHTRYTCVCLASSPAGKRSFSLNVFVSTIAGVS 2578
QY 287 D K T I W 293
Db 2579 DGNPEVTVILNSPSTSLVCEAYSYPATITWFKDGTPLBSNRNIRILPGGRTILOILNAOE 2638
QY 294 D A H TG EA 300
Db 2639 DNAGRYSCVATNEAGEMIKHYEVKVIPIINKGLWGPGLSPKPKVKIKVNNLTLECEA 2698
QY 301 K Q F 304
Db 2699 YAIPSASLSWYKDGQPLKSDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELDF 2758
QY 305 P F H 307
Db 2759 DVNIQVPPSFQKLWEIGNMLDTCRNGEAKDVIINNPI SLYCETNAAPPPTLTWYKDGHP 2818
QY 308 SA P A LDV D 315
Db 2819 TSSDKVLLPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPIIKGANSDL 2878
QY 316 WO SN 319
Db 2879 PBEVTVLNKSALECLSSGPAPRNSWQDGOPLLEDDHHKFLSNGRILQILNTOITDI 2938
QY 320 NT FAS C S 326
Db 2939 GRVYCVAEANTAGSAKKYFNLNVHPSPVIGPKSENLTVVVNNFISLTCEVSGFPDLSW 2998
QY 327 T D MCI 331
Db 2999 LKNEQPIKLTNTLIVPGGRTLQIIRAKVSDGGEYTCIAINQAGESKKFSLTVVYPPSI 3058
QY 332 H V C K L GO 338
Db 3059 KOHDSLSVNVNREGTSVLECESNAVPPVITWYKGRMITESTHVILADGQMLHIK 3118
QY 339 D R P IK TF 345

```
Db 3119 KAESDGTQVVCRAINAGRDDKNFHLNVVPPSIEGPEREVIVETISNPVLTCDATGI 3178
QY 346 -----Q-----G-----H-----T----- 349
Db 3179 PPPTIAWLKNHKKRIENSDSLVRILSGGSKLQIARSQHSNGYTCIASNMEGKAQYYP 3238
QY 350 -----N-E-V-NA-----IKW--D-P--T----- 360
Db 3239 LSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIQLWKDKPIASGETERIRVS 3298
QY 361 --G--NLL-A-S--C--S--D-----D-----D-----WTL----- 373
Db 3299 ANGSTLNIYGALTSDTGKYTCVATNPAGEBEDRIFENLVNVTPTIRGNKDEAEKMLTLDVT 3358
QY 374 --K-----I-W-----SM-----KO--D--NCV----- 384
Db 3359 SINIECRATGTPPPQINLWLNKGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVTVCASNR 3418
QY 385 -----H-----D-----L--Q----- 388
Db 3419 AGVDNKHYNLOVPAPNMDNSMCTEETVLKGSSTSMACITDCTPAPSAWLRDGQPLGL 3478
QY 389 -----Q-----H-N-----KEI 394
Db 3479 DAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFILKLVLEPPHINGSSEHEE 3538
QY 395 -----YT-----I--K--W----- 399
Db 3539 SVIVNNPLELTCIASIPAPKMTWMKGRPLPQTDQVTLGGGEVLRISTAQVEDTGRYT 3598
QY 400 -----SPTG-----P--GT-----N-----N-----N 408
Db 3599 CLASSPAGDDDKKEYLVRVHPNIAGTDEPRDITVLRNQVLTLEKSDAVPPPVITWLRN 3658
QY 409 -----P-----N-----ANLM-LAS--A--S--F----- 420
Db 3659 GERLOATPRVRLISGGRYLQINADLGDNTANYTCVASNIAGKTRSFILTVNVPNNIKGG 3718
QY 421 -----DSTV-----RL-W-D-V-----DR-----G--I----- 433
Db 3719 PQSLVILLNKSTVLEACIAGVPTPRITWRKOGAVLAGNHARYSILENGLHQSAHVTD 3778
QY 434 -----C-----I--H-----TL-----T----- 439
Db 3779 GRYLCMATNAAGTDRRIDLOVHVPPSIAPGPTMTVIVNVQTLTACEATGIPKPSINWR 3838
QY 440 K--H-----Q-E-----P-V-----Y-----S 447
Db 3839 KNGHLLNVQDNQNSYRLSSGSLVILSPSVDATYECTVTNGAGDDKRTVDLTQVPPS 3898
QY 448 VA-----F-S-----P--D----- 453
Db 3899 IADEPTDFLVTKHAPVITCTASGVPPPSIHWTKNGIRLLPRGDGYRILSSGAIEILATQ 3958
QY 454 -----GRY-----L--ASG--S-F 462
Db 3959 LNHAGRYTCVARNAGSAHRHVTLHVHPPVIOQPSELHVILNPIILLPCEATGTPSPF 4018
QY 463 -----X-----D-----KCV----- 466
Db 4019 ITWQKEGINVNTSGRNHVLPSGGLQISRAVREDAGTYMCVAQNPAQTALGKILNVQVP 4078
QY 467 -----H-----I-W----- 469
Db 4079 PVISPHLKEYVIAVDKPIITLSCADGLPPDPITWTKDGRAIVESIRQVRLSSGSLQIAFV 4138
QY 470 -----N-----T-----Q----- 472
Db 4139 QPGDAGHYTCMAANVAGSSSTSKLTVHVPPIRSTEGHYTVNNSQAILPCVADGIPPT 4198
QY 473 -----T-----G-----A-----L-VH- 478
```

RESULT 13

ADJ70089

ID ADJ70089 standard; protein; 5636 AA.

XX

AC ADJ70089;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human heat mitochondrial protein as a therapeutic target SeqID1895.

XX

KW mitochondrial; human; screening assay; diabetes mellitus;

KW Huntington's disease; osteoarthritis;

KW Leber's hereditary optic neuropathy; LHON;

KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;

KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;

KW neuroprotective; neurotropic; antidiabetic; anticonvulsant; antiarthritic;

KW osteopathic; ophthalmological; cytostatic.

XX

OS Homo sapiens.

XX

PN WO2003087768-A2.

XX

PD 23-OCT-2003.

XX

PF 04-APR-2003; 2003WO-US010870.

XX

PR 12-APR-2002; 2002US-0372843P.

PR

PR 17-JUN-2002; 2002US-0389987P.

PR

PR 20-SEP-2002; 2002US-0412418P.

XX

XX (MITO-) MITOKOR.

PA (BUCK-) BUCK INST AGE RES.

XX

PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

PI Warnock DE;

XX

XX WPI; 2003-845369/78.

DR

XX Identifying a mitochondrial target for drug screening assays and for

PT treating diseases associated with altered mitochondrial function,

PT comprises detecting a modified polypeptide in a sample and correlating

PT with the disease.

XX

PS Claim 1; SEQ ID NO 1895; 180pp; English.

XX

XX This invention relates to novel mitochondrial targets that can be used

CC

CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, neurotropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cyostatic activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.
 XX
 SQ Sequence 5636 AA;

Query Match 71.5%; Score 2614.8; DB 7; Length 5636;
 Best Local Similarity 9.5%; Pred. No. 5.2e-31;
 Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S-----I-----SS--D-----EV----- 8
 DB 1 MISWEVVHTVFLFALLYSSLAQDASPOSEIRAEFFEGASTLAFVFDVTGMYDDLQVVI 60
 QY 9 -----NF-LV----- 12
 DB 61 EGASKILETSLKRPKPLFNALVPHDPEIGPVTTITDDPKKQYELRELYVQGGDCPE 120
 QY 13 -----Y-----R--Y-----IQ----- 17
 DB 121 MSGAIKALEISLPGSFIVFTDARSKDYRLTHEVLQLIQKQSQVVFVLTGDCDDRTH 180
 QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
 DB 181 IGYVBEIASTSSGVFHLDKQVNEVLKWEBAQVQSKVHLLSTDLHQAVNTWRIFP 240
 QY 27 -----T-----I-----FG-----I-----K 31
 DB 241 DPSLKEVTVSLGSPSMIEIRNPLGKLIKGFGLHLLNHNHSAKVNKVEPEAGWTVK 300
 QY 32 ---S--H-----I-----S--Q----- 36
 DB 301 TSSSGRHSVRITGLSTIDFRAGFSKPTLDFKKTVSFVQGIPTVYVLLNTSGISTPARID 360
 QY 37 -----S-----NI-----N-----G----- 41
 DB 361 LLELLSISGSSLKTIPVKYYPHRKPYGIMNISDFVPPNEAFFLKVTDYDKDYLFORVSS 420
 QY 42 -----A-----L-----VP-----P-----A 47
 DB 421 VSFSSIVDPAPKVTMPKTEGTYLQPCQIPCSVDLSLLPFTLSFVRNGVTGLGVQYLKESA 480
 QY 48 -----A-----L-----I-----S-----IIQ-----RG----- 56
 DB 481 SVSLDIKAVTLSDGFFVECIASVAGTGRAQTFDVSPPPPVQVPPNVTVTGERAVLT 540
 QY 57 -----L--Q-----L-----Y--V--E----- 61
 DB 541 CLIIAIVDYNLTQWRNDRVRLAEPARTLANLSLELKSFKFNDAEYHCWVSSEGGSS 600
 QY 62 -A-----EVS-----NE----- 68
 DB 601 AASVFLTVQBPVKVMPKQSFSGSEVSMCSATGPKPKIAWTVMDFIVGSHRYRM 660
 QY 69 --DGLTF-----D-----G-----R--P-----I-- 78
 DB 661 TSDGTLPIKNAAPKADAGIYGLASNSAGTDKQNSTLRYIEAPKLMVQSELLVALGDIIV 720
 QY 79 -E--S-----L--S--LI-----DA-----V----- 87
 DB 721 MECKTSGIPPPQVQKFKGDLRLPSTFLIIDPLLLGLLKIQETQDLQDADYTCVAINAAGR 780

QY 88 -----M-----P-DV-----VQ-----T----- 94
 DB 781 ATGKIITLDVSGPPVFIOEPADVSMEIGSNVTLPCYVQGYPEPTIKWRLDNMFISRPFS 840
 QY 95 -----R----- 95
 DB 841 VSSISQLRTGALFILNLWASDKGTICEAENQFKIQSETTVTVTVGLVAPLIGISPSVAN 900
 QY 96 ---QQ-----A-----Y-----R--D--KL-- 103
 DB 901 VIEGQQLTLPCTLLAGNPPIERWIKNSAWMLQNPYITVRSQGLHIERVOLQDGGGYTC 960
 QY 104 ---A-----Q--QQ-----AA----- 109
 DB 961 VASNVAGTNNKTSVVVVHVLPTIHCQQILSTIEGIPVTLPCASGNPKPSVLSKKGEL 1020
 QY 110 ---AAA---A-A-----A--AA--A----- 118
 DB 1021 ISTSAKFSAGADGSLVWSPGESEGYVCTATNTAGYAKRKVQLTVVVRPVFGDLRG 1080
 QY 119 -SQ-----O-----CS----- 123
 DB 1081 LSQDKPVEISVLAGEEVTLPCEVKSUPPIITWAKETQLISPPSRHFLPSSGMKITET 1140
 QY 124 -----A-----K-----N----- 126
 DB 1141 RTSDSGMVLCVATNIAGNVTOAVKLNHVHPKIQRGPKHLKVQVQGVQVDIPCAOQTPLP 1200
 QY 127 -----GE--N--T-----A-----N--G----- 133
 DB 1201 VITWSKGGSTMLVDBHHVSNPDGTLSDQATPSDAGIVTCVATNIAGTDETEITLHVQE 1260
 QY 134 -----E-----E--N-----GA-----H----- 139
 DB 1261 PPTVELEPPYNTTFQERVANQRIEPPCPAKGPKPTIKWLHNGRELGTREPGISILEEG 1320
 QY 140 T--IA-----N-----N--HT-----DM----- 148
 DB 1321 TLLVIAVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKEQVSNVSVLLNQLTNL 1380
 QY 149 -----M-----E----- 150
 DB 1381 FCEVEGTSPPIIMWYKDNVQVTTESITQTVNNGKILKLFATPEDAGRYSCKAINIAGTS 1440
 QY 151 ---VD-----G-----DV--E-----I-----P-----PN 160
 DB 1441 QKYFNIDVLVPPTIIGTNFPKEVSVVLRNDRVALECOVKGTPTPDHFKDGRPLFLGDPN 1500
 QY 161 -----K-A-----V-----V-----VL 165
 DB 1501 VELLDRGQVLHLKNARENDKGRYQCTVSNAGKQAKDKILTIYIPPSIKGGNVTTDISVL 1560
 QY 166 -----RG-----H----- 168
 DB 1561 INSLIKLECKTRGLPMPAITWYKDGQIPIMSSSQALYIDKQYLHIIPRAQVSDSATYCHV 1620
 QY 169 -----E-S-----E-----VFI-----C--A-----W----- 177
 DB 1621 ANVAGTAKSFHVVDVYVPPMIEGNLATPLNKQVVIASHLTLNCAAGNPSPILTWLKDG 1680
 QY 178 -----N-----P-V--SD-- 182
 DB 1681 PVKANDNFRIEAGGKKLEIMSAQIDRGQYICVATSVAGEKEIKYEVDVLPVPAIEGGDE 1740
 QY 183 -----LL-----ASGS-----G-----D-----S-- 191
 DB 1741 TSYFVVMVNNLLELCHVTGSPPTIMLWKGDLIDERDGFKILLNGRKLVIQAQVSNT 1800
 QY 192 -----TA-----R-----I--W 196
 DB 1801 GLYRCMAANTAGDHKKEFEVTVHVPPTIKSSGLSERVWVKYKVALQCIANGIPNSITW 1860
 QY 197 -----NL-----S 199

Db 1861 LKDDQPNTAQGNLKIQSSGRVLQIAKTLLEDAGRYTCVATNAAGETQOQHQLHVHPPS 1920
QY 200 -E-----N-----S-----T-----SGST----- 207
Db 1921 LEDAGKMLNLTVLSNPVQLEKKAAGNPVPVITWYKDNCLLSGSTMTFLNRQOIIDIES 1980
QY 208 -Q-----L-----V-----LR-H-----C-----I-- 215
Db 1981 AQISDAGIYKCAINSAGATELFVSLQVHVAPSISSGNMNAVAVVNNVPVLECEARGIPA 2040
QY 216 -----R-----E-----G-----G--Q--D-- 221
Db 2041 PSITWLKDGSPVFSFNGLQVLSGGRILALTSTQISDGRYTCVAVNAAGEKORDILRV 2100
QY 222 -VPSN-----KD-----VT-----S-L- 231
Db 2101 YVPPNIMEERQNVSVLISQAVELLQSDAIPPTTLTWLKDGHPLLLKPGLSISENRSVLK 2160
QY 232 -----D-----W-----N-----S-----EG--TLL-- 240
Db 2161 IEAQVQDGTGRYTCATNVAGTEKKNNYVNIWPPNIGGSDELTLQTVIEGNLSILCE 2220
QY 241 AT-----GS--Y--D-----G-----F-----A- 249
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRLSGRQLQISIAEKSDAALYSCVASNAVAGTAK 2280
QY 250 -----R-----I-----WTXDG----- 256
Db 2281 KEYNLQVIRPTTNGSGHTEIIVTRGKSISLECEVOGIPPTVTWMDKGHPLIKAGV 2340
QY 257 -----NL-AS-----T-----L-----GOH----- 265
Db 2341 EILDEGHILQKNIHVSIDTGRYCVAVNVAGMTDKYDLSVHAPPSTIGNHRSPENISV 2400
QY 266 -K-----G-PI-----FALK--W----- 274
Db 2401 EKNSVSLTCEASGIPLPSTWTF--KDGWPVLSNSVRILSGGRMLRLMOTTMEDAQYTC 2458
QY 275 -----N-----K-K-----GN----- 279
Db 2459 VRNAAGEERKIFGLSVLPVPHVIGENTLEBVKVKEKQSVTLTCEVTGNPVPBITWHKOG 2518
QY 280 -----F--I-----L-----S--AGV- 286
Db 2519 QPLQDEAHIIISGRFLQITNVQVPHTRYTCCLASSPAGKSRSPSLNVFVSPTIAGVG 2578
QY 287 -D--K--T-----T-----I-W----- 293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPATITWPKDGTPLSNRNIRILPGGRTLQILNAQ 2638
QY 294 -D-----A-----H-----TG-----E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGLSPKEVKIKVNNLTLECE 2698
QY 300 A-----K-----Q-----Q----- 303
Db 2699 AYAI PSASLSWYKDGQPLKSDDDHVNIAANGHTLQIKEAQISDGRYTCVASNIAGEDEL 2758
QY 304 F-----P-F-----H----- 307
Db 2759 FVNIQVPPSFQKLWEIGNMLDTRNGEAKDVIINNPISLYCEETNAAPPPTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTSSDKVILIPGGRVLIQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878
QY 316 -----WO-----SN----- 319
Db 2879 LPEVTVLVNKSALIECLSSGSPAPRNSWQKQGPILLEDDHKFLSNGRILQILNTQITD 2938
QY 320 -----NT-----FAS--C-----S 326

Db 2939 IGRYVCVAENTAGSAKKYFNLNVHVPSPVIGPKSENLTVVNNFISLTCEVSGFPDPDLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKXQPIKLTNTLIVPGRTLQIIRAKVSDGEYTCIAXNAGESKKKFSLTVYVPPS 3058
QY 332 -H-----V-----C-----K-----L--GQ-- 338
Db 3059 IKDHDSESLSVNVNREGTSVLECESNAVPPPVITWYKNGRMITESTHVEILADGQMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KKAESVDTQYVCRAINVAGRDDKNFNLNVYVPPSIEGPEREVIVETISNPVLTCDATG 3178
QY 344 -T-----F-----Q--G--H--T----- 349
Db 3179 IPPPTIAWLKNYKRIENSLSVRILSGSKLOIARSQSDSGNYTCIASNMEGKAQYY 3238
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360
Db 3239 FLSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTLIQWKDKPIASGETERIRV 3298
QY 361 -G--NLL-A--S-----C--S-----D-----D-----MTL-- 373
Db 3299 SANGSTNIYGALTSDTGKYTCVATNPAGEEDRIFNLNVYVPTIRGNKDEAEKLTMYD 3358
QY 374 -K-----I-W-----SM-----KQ--D--NCV-- 384
Db 3359 TSINIECRXTGTPPPQIKNLKNGLPLPLSSHRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418
QY 385 -----H-----D-----L--Q-- 388
Db 3419 RAGVDKNHYNLQVAFAPPNMDSMGTEITVLKSSSTSMACITDGTAPASMAWLDGQPLG 3478
QY 389 -----Q-----H-N-----KE 393
Db 3479 LDAHLTVSTHGMVLQLLKKAETEDSGKYTCIASNEAGEVSKHFLKVLPEPHINGSEHEE 3538
QY 394 I-----YT-----I--K--W----- 399
Db 3539 ISVIVNPLELTCIASGIPAPKMTWMMKGRPLPQTDQVQTLGGGEVLRISTAQVEDTGRY 3598
QY 400 -----SPTG-----P--GT-----N----- 407
Db 3599 TCLASSPAGDDDKXYLVRVHVPPNIAGTDEPRDITVLRNRQVTLCKSDAVPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS-A--S--F----- 420
Db 3659 NGERLOATPRVRLSGGRYLQINNADLGDNTANYTCVASNIAGKTTREFILTVNVPPNIK 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I-- 433
Db 3719 GPQSLVILLNKSTVLECIAGVPTPRITWRKQGAVLAGNHARYSILENGFLHIQSAHVT 3778
QY 434 -----C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCMATNAAGTDRRIDLVHVHPPSIAPGPTMTVIVNVQTLTACEATGPKPSINW 3838
QY 440 -K--H--Q-E-----P-V--Y----- 446
Db 3839 RKNHLLNVNDQNSYRLSSGSLVSIISPSVDDTATYECTVTNGAGDKRTVOLTQVPP 3898
QY 447 SVA-----F-S-----P--D----- 453
Db 3899 STADEPTDLVTKHAPAVITCTASGVPPFSIHWTNGIRLLPRGDGRYRILSSGAIELAT 3958
QY 454 -----GRY-----L--S--S-- 461
Db 3959 QLNHAGRYTCVARNAGSAHRHVTLVHPPVPIQOPQSELHVLNPNILLPCEATGTPSP 4018
QY 462 F-----D-----KCV----- 466
Db 4019 FITWQKEGINVNTSGRNHVLPSGGIQLISRAVREDAGTYMCVAQNPAQTALGKILNVQV 4078

QY 467 -----H-----I-W----- 469
 Db 4079 PPVISPLKKEYIVADKPITLSCADGLPPDITWKKDGRAIVESIRQVLSSGSLQIAF 4138
 QY 470 -----N-----T-----Q----- 472
 Db 4139 VQPGDAGHYTCMAANVAGSSSTSKLTIVHVPPIRSTEGHYTVNENSAQLPCVADGPT 4198
 QY 473 -----T-----G-----A-----L-VH 478
 Db 4199 PAINWKKDNVLLANLLGKYTAEPYGELILENVLEDSGFYTCVANNAAGDTHVSLTVH 4258
 QY 479 -----S-----Y-----R 481
 Db 4259 VLPTFTPLPGDVSLNKGEQLRLSKATGIPLPKLTWTFNNNIIPAHFDSVNGHSELVIER 4318
 QY 482 -----GT-----G-----E-----V-C----- 490
 Db 4319 VSKEDSGTYVCTAENSFGVKATGFFVYKPPVFKGDPYPSNWIEPLGNAILNCEVKGDP 4378
 QY 491 -----WN-----A-----AGD----- 496
 Db 4379 TPTIQNRKGVDIIEISHRIQLGSLAIYCTVNEDAGDYTCVATNEAGVVERSMSTLQ 4438
 QY 497 -----K-V-----A-----SAS----- 503
 Db 4439 SPPIITLEPVETVINAGKILNCOATGEPOPTITWSRQHSISWDDRVNVLNNSLYIA 4498
 QY 504 -----D-----GSV-----C-V----- 509
 Db 4499 DAQKEDTSFEFCVARNLWGLVLRVPVIVQVHGFSQMSAWRACSVTCGKGIQKRSRLCN 4558
 QY 510 --L-----DL--R-----K 514
 Db 4559 QPLPANGGKPCQGSDLERNCONK 4582

RESULT 14
 ADJ83137
 ID ADJ83137 standard; protein; 5636 AA.
 AC ADJ83137;
 AC AC
 DT 06-MAY-2004 (first entry)
 XX Human hemocentin protein - SEQ ID 128.
 XX NOVX; cytostatic; antiarteriosclerotic; antidiabetic; antiasthmatic;
 KW antiasthmatic; antiinflammatory; respiratory; antiarthritic;
 KW dermatological; antibacterial; cerebroprotective; vasotropic; cardiant;
 KW haemostatic; hypotensive; hepatotropic; neuroprotective; anorectic;
 KW neurotropic; antitumor; muscular; immunosuppressive; gynaecological;
 KW antiparkinsonian; anticonvulsant; tranquiliser; osteopathic;
 KW antifertility; antipain; cardiomyopathy; atherosclerosis; diabetes;
 KW cell signal processing; metabolic pathway; asthma; allergy; emphysema;
 KW autoimmune; graft-versus-host; arthritis; cancer; stroke; haemophilia;
 KW obesity; Alzheimer's; pain; chromosome mapping; tissue typing; human;
 hemocentin.
 XX Homo sapiens.
 XX US2003170630-A1.
 XX 11-SEP-2003.
 XX 21-DEC-2001; 2001US-00032189.
 XX 21-DEC-2000; 2000US-0257495P.
 XX 22-DEC-2000; 2000US-0258171P.
 XX 20-FEB-2001; 2001US-0269940P.
 XX 08-MAR-2001; 2001US-0274192P.

PR 22-MAR-2001; 2001US-0277826P.
 PR 29-MAR-2001; 2001US-0279840P.
 PR 11-APR-2001; 2001US-0282981P.
 PR 13-APR-2001; 2001US-0283656P.
 PR 31-JUL-2001; 2001US-0309247P.
 PR 10-AUG-2001; 2001US-0311754P.
 PR 17-AUG-2001; 2001US-0313331P.
 XX (ALSO/) ALSOBROOK J P.
 PA (TCHE/) TCHERNEV V T.
 PA (LIUX/) LIU X.
 PA (SPYT/) SPYTEK K A.
 PA (ZERH/) ZERHUSEN B D.
 PA (PAT7/) PATTURAJAN M.
 PA (LEPL/) LEPLEY D M.
 PA (BURG/) BURGESS C E.
 PA (SHIM/) SHIMKETS R A.
 PA (GROS/) GROSSE W M.
 PA (SZEK/) SZEKERES E S.
 PA (VERN/) VERNET C A M.
 PA (LILL/) LI L.
 PA (CASM/) CASMAN S J.
 PA (BOLD/) BOLDOG F L.
 PA (GORM/) GORMAN L.
 PA (GANG/) GANGOLLI E A.
 PA (FERN/) FERNANDES E R.
 PA (RIEG/) RIEGER D K.
 PA (EDIN/) EDINGER S R.
 PA (GUNT/) GUNTHER E.
 PA (MILL/) MILLET I.
 PA (SCIO/) SCIORE P.
 PA (ELLE/) ELLERMAN K.
 PA (MACD/) MACDOUGALL J R.
 PA (SMIT/) SMITHSON G.
 XX
 PI Alsobrook JP, Tchernev VT, Liu X, Spytek KA, Zerhusen BD;
 PI Patturajan M, Lepley DM, Burgess CE, Shimkets RA, Grosse WM;
 PI Szekeres ES, Vernet CAM, Li L, Casman SJ, Boldog FL, Gorman L;
 PI Gangolli EA, Fernandes ER, Rieger DK, Edinger SR, Gunther E;
 PI Millet I, Sciore P, Ellerman K, Macdougall JR, Smithson G;
 XX WPI; 2003-898249/82.
 DR New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,
 XX preventing or treating NOVX-associated polypeptide disorder, e.g.
 PT cardiomyopathy, atherosclerosis, diabetes, cancer, Parkinson's disease or
 PT asthma.
 XX Disclosure; SEQ ID NO 128; 263pp; English.
 PS The invention relates to a novel isolated NOVX polypeptide. The
 XX polypeptide demonstrates cytostatic, antiarteriosclerotic, antidiabetic,
 CC antiasthmatic, antiinflammatory, antiparkinsonian, anticonvulsant,
 CC antiasthmatic, antiasthmatic, antiinflammatory, respiratory,
 CC antiarthritic, dermatological, antibacterial, cerebroprotective,
 CC vasotropic, cardiant, haemostatic, hypotensive, hepatotropic,
 CC neuroprotective, anorectic, neurotropic, antitumor, muscular,
 CC immunosuppressive, gynaecological, antiparkinsonian, anticonvulsant,
 CC ophthalmological, osteopathic, antiparkinsonian, anticonvulsant,
 CC tranquiliser, analgesic, nephrotropic, antinfertility and antilipaemic
 CC activities. The NOVX polypeptide, nucleic acid or antibody of the
 CC invention may be useful for treating or preventing a NOVX-associated
 CC disorder, such as cardiomyopathy, atherosclerosis, diabetes or a disorder
 CC related to cell signal processing and metabolic pathway modulation.
 CC Furthermore, the NOVX polypeptides may be useful for diagnosing, treating
 CC or preventing diseases such as asthma, allergies, emphysema, autoimmune
 CC disease, graft-versus-host disease, arthritis, cancer, stroke,
 CC haemophilia, obesity, Alzheimer's disease and pain. The nucleic acids may
 CC be used as hybridisation probes, in chromosome mapping, tissue typing,
 CC preventive medicine or pharmacogenomics. The current sequence is that of
 CC a protein of the invention which is related to human NOVX protein.
 XX Sequence 5636 AA;
 SQ

Query Match 71.5%; Score 2614.8; DB 7; Length 5636; Best Local Similarity 9.5%; Pred. No. 5.2e-31; Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;			
QY	1 M-S-----I-SS--D-----EV-----	8	
Db	1 MISWEVHTVFLFALLYSSLAQDASPOSEIRAEFFEGASTLAFVFDVTGSMYDDLQVVI	60	
QY	9 -----NF-LV-----	12	
Db	61 EGASKILETSKRPKPLNFALVPHPDPEIGPVITTTDPKFKQYBELRELYVGGDCPE	120	
QY	13 -----Y-----R--Y-----LQ-----	17	
Db	121 MSIGAIKIALEISLPGSFIVFTDARKQYRLTHEVLQLIQKQSQVFWLJTGDCDRTH	180	
QY	18 -----E--SG--F-----S--H--S-----A-----F 26		
Db	181 IGYVVEEIASTSSGVFHLDDKQVNEVLKWVEAVQAKVHLLSTDHLEQAVNTWRIPP	240	
QY	27 -----T-----FG--I-----K 31		
Db	241 DPSLKVTVSLGSPSMIEIRNPLGKLIKGFGLHLLNHSKVVNVKEPEAGHWTK	300	
QY	32 ---S--H-----I-----S--Q-----	36	
Db	301 TSSSGRHVRITGLSTIDFRAGFRKPTLDFKKTVSFVQGIPTYVLLNTSGISTPARID	360	
QY	37 -----S-----NI-----N-----G-----41		
Db	361 LLELLSIGSSSLKTI PVKYYPHRKPYGIWNISDFVPPNEAFFLKVTGYDKDDYLFQVSS	420	
QY	42 -----A-----L--VP-----P-----A 47		
Db	421 VSPSSIVDPAPKVTMEKTPGYLQPGQIPCSVDLSLLPFTLSFRNGVTGLVDQYLKESA	480	
QY	48 -----A--L-----I-----S--IIQ-----KG-----56		
Db	481 SVSLDIKVTLSDEGFVECIASVAGTGRAQTFDVSERPPVIOQVNNVTVTGERAVLT	540	
QY	57 -----L--Q-----P--V--E-----61		
Db	541 CLIIISADVNLTORNDROVRLAEPARTLANLSLELKSVEKFNDAEGEYCHMVSSEGGSS	600	
QY	62 -A-----EVS-----NE-----68		
Db	601 AASVFLTVQEPKVTMPKQKSTGGSEVSMCSATGYPKPKIAWTNDMFIVGSHRYRM	660	
QY	69 --DGTLF-----D-----G-----R--P-----I--78		
Db	661 TSDGTLFIKNAAPKDAGIYGLASNSAGTDKQNSTLYRIEAPKLMVVQSELLVALGDITV	720	
QY	79 -E--S-----L--S--LI-----DA-----V-----87		
Db	721 MECKTSGIPPPQVWFKGDLELPSTLIIDPLGLLKIQETQDLBAGDYTCVAINAAGR	780	
QY	88 -----M-----P-DV-----VQ-----T-----94		
Db	781 ATGKITLDVGGSPPVFIOEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMPFISRPFS	840	
QY	95 -----R-----95		
Db	841 VSSISQRTGALFILNWDKTYICEAENQFKIQSETTVTVTGLVAPLIGISPSVAN	900	
QY	96 ---QQ-----A-----Y-----R--D--KL--103		
Db	901 VIEGQQLTLPCTLLAGNPIPERRWIKNSAMLLQNPYITVRSGLSHIERVOLQDGEYTC	960	
QY	104 ---A-----A-----Q--Q-----AA-----109		
Db	961 VASNAGTNNKTTVVVVHVLPTTQHQQQLLSTIEGIPVTLPCKASGNPKPSVIMSKGEL	1020	
QY	110 ---AAA--A-A-----A--AA--A-----118		
Db	1021 ISTSSAKFAGAGDGLYVVSPEGESGEYVCTATNTAGYAKRKVQLTVYVPRVFGDLRG	1080	
QY	119 -SQ-----O-----GS-----123		
Db	1081 LSODKPEVSIAGEEVTLPCEVKSUPPIITWAKETQLISPPSPRHTFLPSSGMKITET	1140	
QY	124 -----A-----K-----N-----126		
Db	1141 RTSDSGMYLCVATNIAGNVTAQVKNLVNHPVKIQRGPKHLKVQVGQVVDIPCAQGTPLP	1200	
QY	127 -----GE--N--T-----A-----N--G-----133		
Db	1201 VITWSKGGSTMLVDEGHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDEBITLHVOE	1260	
QY	134 -----E--N-----E--N-----GA-----H-----139		
Db	1261 PPTVEDEPPYNTTFOERVANQRIEPPCPAKGPKPTIKWLHNGRELGTREPGISILEEG	1320	
QY	140 T--IA-----N-----N--HT--DM-----148		
Db	1321 TLLVIASTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKDEQVSNVSVLLNQLTNL	1380	
QY	149 -----M-----E-----150		
Db	1381 FCEVEGTSPSIIMWKDQNVQVTTESSTIQTVNNGKILKLFATPEDAGRYSCKAINIAGTS	1440	
QY	151 ---VD-----G-----DV--E-----I-----P--PN 160		
Db	1441 QKYFNIDVLVPPTIIGTNFPKEVSVVLRDVALECOVKGTPTPDHFKDQKFLGDPN	1500	
QY	161 -----K--A-----V-----VL 165		
Db	1501 VELLDRGVHLKGNARRNDKGRYQCTVSNAGQAKDIKLTIVIPSIKGNVTTDISVL	1560	
QY	166 -----RG-----H--168		
Db	1561 INSLIKLECKTRGLPMPAITWYKQGPIMSSSQALYIDKQYLHI PRAQVSDSATYCHV	1620	
QY	169 -----E--S-----E-----VFI-----C--A-----W-----177		
Db	1621 ANVAGTAESFHVVDVVPPIEGNLATPLNKQVIAHSITLECNAAAGNPSPIITLWKDGV	1680	
QY	178 -----N-----P--V--SD--182		
Db	1681 PVKANDNFRIEAGKKLEIMSAQEDIRGOYICVATSVAGEKEIKYEVDVLVPPAIEGDE	1740	
QY	183 -----LL-----ASGS-----G--D-----S--191		
Db	1741 TSVFIVMNNLLELDCHVTGSPPTIMLWKDGLIDERDGFKILLNGRKLVIQAQVSNV	1800	
QY	192 -----TA-----R-----I--W 196		
Db	1801 GLYRCHMAANTAGDHKKEFEVTVHPPTIKSSGLSERVVVYKYPVALQCIANGIPNSITW	1860	
QY	197 -----NL-----S-----199		
Db	1861 LKDDQPVNTAQNKLQSSGRVLIQIAKLTLEDAGRYTCVATNAAGTQOQHILHVHEPPS	1920	
QY	200 -E-----N--S-----T-----SGST-----207		
Db	1921 LEDAGKMLNETVLVSNPQLECKAAGNPVPVITWYKDNCLLSGTSMTFLNRRGIIDIES	1980	
QY	208 -Q-----L--V-----LR--H-----C-----I--215		
Db	1981 AQISDAGIYKCVAINASAGATELFYSLQVHVAPVPSIGSNMNVAVVNVNVRLECEARGIPA	2040	
QY	216 -----R-----E--G-----G--Q--D--221		
Db	2041 PSLTMLKQSPVSSFNGLQVLSGGRILALSTQISDTRYTCVAVNAAGKORDDLRV	2100	
QY	222 -VPSN-----RD-----VT--S--L--231		

Db 2101 YVPPNIMGEQNSVLISQAVELLCSQDAIPPTLTWLDKGHPLKKPGLSISENRRSVLK 2160
QY 232 -D- - - - -W- -N- -S- - - - -EG- -TLL- - 240
Db 2161 IEDAQVDTGRYTCEATNNAVKTEKKNVNNWPPNIGSGDELQTLTVIEGNLISLCE 2220
QY 241 AT- - - - -GS- -Y- -D- - - - -G- - - - -F- - - - -A- 249
Db 2221 SSGIPPPNLIWKKGSPVLDTMGRVRLISGRQIQISIAEKSDAALYSVASNVAGTAK 2280
QY 250 - - - - -R- - - - -I- - - - -WTGQ- - - - - 256
Db 2281 KEYNLQVIRPTITNSGHPTEIIVTRGKSISLCEVQGIPPTVTWKKDCHPLIKAKGV 2340
QY 257 - - - - -NL- -AS- - - - -T- - - - -L- - - - -GOH- - - - - 265
Db 2341 EILDEGHILQKNLHVSQVSDTGRYCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K- - - - -G- -PI- - - - -FALK- - -W- - - - - 274
Db 2401 EKNSVSLTCEASGIPLSTTW- - - - -KDGMPVSLNSVRILSGRMLRLMQTTMEDAGQYTC 2458
QY 275 - - - - -N- - - - -K- -K- - - - -GN- - - - - 279
Db 2459 VVRNAGEERKIFGLSVLPPHIVGENTLEDVKVKERQSVTLTCEVTGNVPPEITWHKOG 2518
QY 280 - - - - -F- -I- - - - -L- - - - -S- - - - -AGV- 286
Db 2519 QPLOEDBAHHIISGRFLQINVOVPHTGRTYTCCLASSPAGHKRSFSLNVFVSPPTAGVG 2578
QY 287 -D- - - - -K- -T- - - - -I- -W- - - - - 293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLFSNRNIRILPGGRTLQILNAQ 2638
QY 294 -D- - - - -A- - - - -H- - - - -TG- - - - -E 299
Db 2639 EDNAGRYSCVATNEBAGEMIKHYEVKVYIPPIINKGLWGLSPKEVKIKVNNLTILECE 2698
QY 300 A- - - - -K- - - - -Q- - - - -Q- - - - - 303
Db 2699 AYALPSASLSWYKDGQPLKDDHVNIAANGHTLOIKEAQISDTGRYTCVASNIAGEDEL 2758
QY 304 F- - - - -P- -F- - - - -H- - - - -H- 307
Db 2759 FVNIQVPPSFQKLWEIGNMLDTGRNGEAKDVINNPFISLYCETNAAPPPTLTWYKDGHP 2818
QY 308 -SA- - - - -P- - - - -A- - - - -LDV- - - - -D 315
Db 2819 LTSSDKVLLPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878
QY 316 - - - - -WO- - - - -SN- - - - - 319
Db 2879 LPBEVTVLNKSAIECLSSGSPAPRNSWQKQGLLEDHDFKFLSNGRILQILNTQITD 2938
QY 320 - - - - -NT- - - - -PAS- -C- - - - -S 326
Db 2939 IGRYCVVAENTAGSAKKYFNFLNVHVPVSGPKSENLTVVVNNFISLTCEVSGFPFPPDLS 2998
QY 327 - - - - -T- - - - -D- - - - -MCI- - - - - 331
Db 2999 WLKNXQPIKLTNTLIVPGGRTLQIRAKVSDGGEYTCIAXNAGESKKFSLTVVYVPPS 3058
QY 332 -H- - - - -V- -C- - - -K- - - - -L- -GO- - - 338
Db 3059 IKHDSLSLVNVRECTSVSLECSNAVPPPTVWYKNGRMITESTHVEILADGQWLHI 3118
QY 339 -D- - - - -R- - - - -P- -IK- - - - - 343
Db 3119 KKAESVDTGOYVCRAINAGRDKNFHLNVYVPPSIEGPEREIVETISNPFVTLTCDATG 3178
QY 344 -T- - - - -F- - - - -Q- - - - -G- - - - -H- - - - -T- - - - - 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGKLQIARSQHSQSDSGNYTCIASNMEGKAQKY 3238

QY 350 - - - - -N- -E- -VA- - - - -IKW- -D- -P- - - - -T- - - - - 360
Db 3239 FLISQVPPSVAGAEIPSDSVLLGENVELVNCANGIPTPLIQMLKDGKPIASGETERIRV 3298
QY 361 - - - - -G- - -NLL- -A- -S- - - - -C- -S- - - - -D- - - - -D- - - - -MTL- - - 373
Db 3299 SANGSLINIYGALTSITGKYTCVATNPAGEEDRIFNLVVYVPTIRGNKDEAKELMTYVD 3358
QY 374 - - - - -K- - - - -I- -W- - - - -SM- - - - -KO- -D- - - - -NCV- - - 384
Db 3359 TSINIECRXTGTPPPQINMLKGLPLPLSSHIRLLAAGQIVIRVRAQVSDVAVYTCASN 3418
QY 385 - - - - -H- - - - -D- - - - -L- - - - -L- - - - -O- - - - - 388
Db 3419 RAGVDNKNHYNLQVAPPNMDNSMGTEITVLKSGSTSMACITDGTAPASMAWLRDQPLG 3478
QY 389 - - - - -Q- - - - -Q- - - - -H- -N- - - - -KE 393
Db 3479 LDAHLLTVSTHGMVLQLLKAFTEDSGKYTCIASNEAGEVSKHFLKLVLEPPHINGSHEE 3538
QY 394 I- - - - -YT- - - - -I- -K- -W- - - - - 399
Db 3539 ISVINNPLELTCIASGIPAPKMTWMDGRPLPQTDQVOTLGGGEVLRISTAQVEDTRY 3598
QY 400 - - - - -SPTG- - - - -P- - - - -GT- - - - -N- - - - - 407
Db 3599 TCLASSPAGDDKEYLVRVHVPPNIAGTDEPRDITVLRNRQVLTLECKSDAAPPVITWLR 3658
QY 408 N- - - - -P- - - - -N- - - - -ANLM- -LAS- -A- -S- -F- - - - - 420
Db 3659 NGERLQATPRVRLSGRYLQINNADLGDANTVYTCVASNIAGTKTREFILTVNVPNIK 3718
QY 421 - - - - -DSTV- - - - -RL- -W- -D- -V- - - - -DR- - - - -G- -I- - - - - 433
Db 3719 GPQSLVILLNKSTVLECIABGVPTPRITWRKDGAVLAGNHARYSILENGLHIQSAHVD 3778
QY 434 - - - - -C- - - - -I- - - - -H- - - - -TL- - - - - 439
Db 3779 TGYLCMATNAAGTDRRIDLQVHVPPSIAPGTNTMTVIVNVQTTLACEATGIPKPSINW 3838
QY 440 -K- -H- - - - -Q- -E- - - - -P- -V- - - - -Y- - - - - 446
Db 3839 RKNHLLNVQDQNSYRLSSGSLVILSPSVDATYECTVTNGAGDDKRTVDLTVOQVPP 3898
QY 447 SVA- - - - -F- -S- - - - -P- - - - -D- - - - - 453
Db 3899 STADEPTDFLVTKHAPAVITCTASGVPPFSIHWTNGIRLLPRGDGYRILSSGAIFILAT 3958
QY 454 - - - - -GRY- - - - -I- - - - -L- - - - -ASG- -S- - - - - 461
Db 3959 QLNHAGRYTCVARNAGSAHRHVLHVHPPVIOQPSELHVLNNPILLPCEATGTPSP 4018
QY 462 F- - - - -D- - - - -D- - - - -KCV- - - - - 466
Db 4019 FITWQEGINVTSGRNHVLPSGLQISRAVREDAGTVMCAQNPAGTALGKILNVQV 4078
QY 467 - - - - -H- - - - -I- -W- - - - - 469
Db 4079 PPVISPHLKEYVIADVDPKLTILSCADGLPPPDIITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 470 - - - - -N- - - - -T- - - - -Q- - - - - 472
Db 4139 VQPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTVNENSOAILPCVADGIFT 4198
QY 473 - - - - -T- - - - -G- - - - -A- - - - -L- -VH 478
Db 4199 PAINWKNVLLANILGKYTABPYGELILENVLEDGFPYTCVANNAGEDTHTVSLTVH 4258
QY 479 - - - - -S- - - - -Y- - - - -R 481
Db 4259 VLPTTELPDGLNKGQRLRUSCKATGIPKLTWTFNNIIPAHFDSVNGHSELVIER 4318

95 -----R----- 95
841 VSSISQRTGALFILNLWASDKGTIYCEAENQFKIOSETTVTGLVAPLIGISPSVAN 900
96 ---QO---A---Y---R---D---KL--- 103
901 VIEGQQLTLPCTLLAGNPPIERRWIKNSAMLLQNPYITVRSQSLHIERVQLQDGGGYTC 960
104 ---A---Q---QO---AA--- 109
961 VASNVAGTNNKTSVVVHVLFTHQGOQILSTIBGIPVTLFCKASGNPKFVIMSKKGEL 1020
110 ---AAA---A---A---A---A--- 118
1021 ISTSSAKFAGADGSLYVVSPEGESEGYVCTATNAGYAKRVQLTVYVRVFGDLRG 1080
119 SQ-----Q-----GS----- 123
1081 LSQDKPVEISVLAGEEVTLPCEVKSLLPPIITMAKETQLISPFSPRHTFLPSSGSMKITET 1140
124 ---A---K-----N----- 126
1141 RTSQGYLCVATNIAGNVTQAVKLVNHPVKIQRGPKHLKVQGVQDVIPCAQGTPLP 1200
127 ---GE---N---T---A---N---G----- 133
1201 VITWSKGGTNLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260
134 ---E---N---GA---H----- 139
1261 PPTVEDLEPPYNTTFQERVANQRIEFPCKAKGTPKPTIKWLHNGRELTGREPGISILEEG 1320
140 T---IA---N---N---HT---DM----- 148
1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKQKEQVSNVSVLLQNLNL 1380
149 ---M-----E----- 150
1381 FCEVEGTPSPIMWYKDNQVTSSTIQTVNNKILKLFATPEDAGRYSCKAINAGTS 1440
151 ---VD---G-----DV---E-----I---P-----PN 160
1441 QKVFNDVLVPPTIIGTFNFKVSNLNRDVALEQVKGTGPPDIHWFKDGKFLFLGDPN 1500
161 ---K---K---A-----V-----VL 165
1501 VELLDRQVLHLKNARNDKGRYOCTVSNAGKQAKDIKLTIVIPPSIKGGNVTTDISVL 1560
166 ---RG-----H----- 168
1561 INSLIKLECKTRGLPMPAITWYKDGQIPMSSQALYIDKQYLHI PRAQVSDSATYTCV 1620
169 ---E---S-----E-----VFI---C---A-----W----- 177
1621 ANVAGTAEKSFHDVYVPPMIEGNLATPLNKQVVIASHLSLTCNAAGNPSPIITLWKDGV 1680
178 ---N-----P---V---SD----- 182
1681 PVKANDNFRIEAGKKLEIMSAQEBIDRGQYICVATSVAGEKEIKYEVVDVLVPAIEGGDE 1740
183 ---LL---ASGS-----G---D-----S--- 191
1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKDGQLIDERDGFKILLNKRKLVIQAQVSN 1800
192 ---TA-----R-----I---W 196
1801 GLYRCMAANTAGHKHEFVTVHPPTIKSGLSERVVVVKYKFPVALQIANGIPNPSITW 1860
197 ---NL-----S 199
1861 LKDDQPVNTAGNKKIOSSGRVLIQIAKTILLEDAGRYTCVATNAAGETQQHIQLHVHPPS 1920
200 -E-----N-----S-----T-----SGST----- 207

1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSGSTMTFLNRQIIDIES 1980
208 -Q-----L---V-----LR-H-----C---I--- 215
1981 AQISDAGIYKVAINSAGATELFYSLQVHVAPSIGSNMNVAVVWNNPVRLCEARGIPA 2040
216 -----R-----E---G-----G---Q---D--- 221
2041 PSUTWLKDGSPVSSFNGQLVLSGGRILALATSTQISDTGRYTCVAVNAAGEKORDIDLRV 2100
222 -VPSN-----KO-----VT---S-L- 231
2101 YVPFNMGBEONVSVLISOAVELLQSDAIPPTTLTWLKDGHLLKPKGLSISENSVLK 2160
232 -----D-----W---N---S-----EG---TLL--- 240
2161 IEDAQVQDGTGRYTCEATNVAGTEKKNVNIWVPPENIGGSDLTQLTVEGNLISLCE 2220
241 AT-----GS---Y---D-----G-----F-----A- 249
2221 SSGIPPNLIWKKKGSFVLTDMSGRVRIILSGGRQLOISIAEKSDAALYSCVASNAVGTAK 2280
250 -----R-----I-----WKDG----- 256
2281 KEYNLQVIRPTITNSGSHPTIIVTRGKSISLECEVOGIPPTVTWMDKDHLLIKAKGV 2340
257 -----NL-AS-----T---L-----GOH----- 265
2341 EILDEGHILQKNIHVSDTGRYVCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400
266 -K-----G---PI---FALK---W----- 274
2401 EKNSVLTCEASGIPLPSTTF--KDGWPVLSNSVRILSGGRMLRMQTTMEDAQOYTC 2458
275 -----N-----K---K-----GN----- 279
2459 VVUNAAGEBKIFGLSVLPPIHVGENTLEDVKVKEQSVTLTCEVTGNPVPPIETWHKOG 2518
280 -----F---I-----L---S---AGV- 286
2519 QPLQDEAHIIISGGRFLOITNVQVPHTRGYTCLASSPAGHKRSFSLNVFVSPTIAGVG 2578
287 -D---K---T---T---I---I---W----- 293
2579 SDGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLESNRNIRILPGRTLOILNAQ 2638
294 -D-----A---H-----TG-----E 299
2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGLSPKEVKIKVNNLTLECE 2698
300 A-----K-----Q---Q----- 303
2699 AVAIPASLSWYKDGQPLKSDDHVNIAANGHTLIQKEAQISDTGRYTCVASNIAGEDEL 2758
304 F-----P---F-----H- 307
2759 FVNIQVPSFQKLWEIGMNLDTGRNGEAKDVIINNPIISLYCETNAAPPTLTWYKDGHP 2818
308 --SA-----P-----A-----LDV-----D 315
2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878
316 -----WQ-----SN----- 319
2879 LPBEVTVLNKSALIECLSSGSPAPRNSWQKQDQPLEDDHHKFLSNGRILQILNTQITD 2938
320 -----NT-----FAS---C-----S 326
2939 IGRYVCVAENTAGSAXKYFNLVNHPVPSVIGPKSENLTVVNNFISLTCEVSGFPFPPDLS 2998
327 -----T-----D-----MCI----- 331

Db 2999 WLKXQPIKLTNTLTVPGRTLQIIRAKVSDGGEYTCIAINXAGBSKKFSLTVVVPSS 3058
QY 332 ---H---V---C---K---L---GQ--- 338
Db 3059 IKHDSLSLVNVRGTSVLECESNAVPPVITWYKNGRMITESTHVEILLADGQMLHI 3118
QY 339 ---D---R---P-IK--- 343
Db 3119 KKAESDTCQYVCRAINAGRDDKNFHLNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 ---T---F---Q---G---H---T--- 349
Db 3179 IPPPTIAWLKNYKRIENSOSLEVRILSGGSKLIARSQSDSGNYTCIASNMEGKAQYY 3238
QY 350 ---G---N-E-V-NA---IKW---D---P---T--- 360
Db 3239 FLSIQVPPSVAGAEIPSDSVLGENVELVCNANGIPTPLIQWLKDGKPIASGETERIRV 3298
QY 361 ---G---NLL-A-S---C-S---D---D---MTL--- 373
Db 3299 SANGSTLNIYALTSDTKYTCVATNPAGEBEDRIFNLNVVYPTIRGNKDEAEKLMYVD 3358
QY 374 ---K---I-W---SM---KQ-D---NCV--- 384
Db 3359 TSINIECRXTGTPPPQINMLKNGLPLPLSGSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418
QY 385 ---H---D---D---L---Q--- 388
Db 3419 RAGVDNKNYLVQFAPNMDNSMGTEITVLKGSSTSMACITDGTAPSMAWLRDQPLG 3478
QY 389 ---Q---H-N---KE 393
Db 3479 LDAHVTSTHGMVLQLLKABEDSGKYTCIASNEAGEVSKHFLKVLPEPHINGSBEHEE 3538
QY 394 I---YT---I-K-W--- 399
Db 3539 ISVIVNPLELTCIASGIPAPKWTWMDKGRPLPQDQVQLGGGEVLRISTAQVEDTGRY 3598
QY 400 ---SPTG---P---GT---N--- 407
Db 3599 TCLASSPAGDDDKXYLVRVVPPNIAGTDEPRDITVLNRQVTLCKSDAVPPPVITWLR 3658
QY 408 N---P---N---ANLM-LAS-A---S-F--- 420
Db 3659 NGERLOATPRVIRILSGRYLQINNADLGDPTANYTCVASNIAGKTTREFILTVNVPNIK 3718
QY 421 ---DSTV---RL-W-D-V---DR---G-I--- 433
Db 3719 GPQSLVILLNKSTVLECIAEGVPTPRITWRKQGAVLGNHARYSILENGFLHIQSAHVTD 3778
QY 434 ---C---I---H---TL---T--- 439
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHVPSPAPGNTMTVIVNVQTLACEATGPKPSINW 3838
QY 440 -K-H---Q-E---P-V---Y--- 446
Db 3839 RKNHLLNVQDNQNSYRLLSSGSLVIPSVDATAYECTVTNGAGDKRTVDLTQVQVP 3898
QY 447 SVA---P-S---P-D--- 453
Db 3899 SIADEPTDFLVTKHAPAVITCTASGVFPFPIHWTNGIRLLPRGDGVIRLLSSGAIBLAT 3958
QY 454 ---GRY---L---ASG-S- 461
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIOQOPSELHVLINNPILLPCEATGTPSP 4018
QY 462 F---D---KCV--- 466
Db 4019 FITWQKEGINVTSGRNHAVLPSGGLQISRAVREDAGTYMCVAQNPAQTALGKILNVQV 4078
QY 467 ---H---I-W--- 469
Db 4079 PPVISPHLKEYIAVDKPIITLSCEADGLPPPDITWHKDGRAIVESIRQVRVLSGSLQIAF 4138

QY 470 ---N---T---Q--- 472
Db 4139 VQPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRIRSTEGHYTWNENSQAILPCVADGPT 4198
QY 473 ---T---G---A---L-VH 478
Db 4199 PAINKKONVLLANLIGKYTAEPYGELILENVVLEDSGPYTCVANNAGEDTHVSLTVH 4258
QY 479 ---S---Y---R 481
Db 4259 VLPFTTELPDVSLLNKGEQLRLSCKATGIPLKLTTWPNNNIIPAHFDSVNGHSELVIER 4318
QY 482 ---GT---G---G---IF---E---V-C--- 490
Db 4319 VSKEDSGTYVCTAENSUGFVKATGFVYVKEPPVFKGDPYPSNWIEPLGGNAILNCEVKGDP 4378
QY 491 ---WN---A---AGD--- 496
Db 4379 TPTIQNRKGVDEIESHRIROLGNSLAIYGTVNEDAGDYTCVATNEAGVVERSMSTLQ 4438
QY 497 ---K-V---G---A---SAS--- 503
Db 4439 SPPIITLEPVETVINAGGKIILNCQATGEPQPTITWSROGHSISWDDRVNVLNNSLYIA 4498
QY 504 ---D---GSV---C-V--- 509
Db 4499 DAQKEDTSEFECVARNLMSVLVRVPVIVQVHGGFQMSAWRACSVTCGKGIOKRSRLCN 4558
QY 510 ---L---DL---R---K 514
Db 4559 QPLPANGGKPCQSGDLEMENCONK 4582

Search completed: January 3, 2005, 15:19:37
Job time : 119.333 secs

This Page Blank (uspto)

Matches 434; Conservative 61; Mismatches 12; Indels 3805; Gaps 385;

1 MS-----SSDE-----V-----N-F-L-----11
88 MSRLCNGVQCMDSDEGPHRELGQNCRLGQHCHVPTLDGFTCYCNSSFQADGKT 147
12 VY-R-----Y-LQ-----E-----18
148 CKDFDESVVGTCSQLCTNTDGSFICGCVGYLLQPNRSCAKNEPDRPPVILLANSQ 207
19 -----SG-----FS-----H-SA-----F 26
208 NILATYLSGAQSVITPTSTQTTAMDFSVANETVCWVHVGDSAAQATOLKCARMPGLKGF 267
27 -----TFGIK-S-HISO-----S-N-----I-----NG-----41
268 VDEHTINISLHVEQMAIDMLTGNFYFVDDIDDRFVCRNRGTCVTLLDLELYNPKG 327
42 -AL-----VP-----P-----A-A-48
328 IALDPAMKVFTDYGOIPKVERCDMDQNRKLVDSKIVFPHGITLDLYSLRYWADAY 387
49 -LI-----SIIQK-L-Q-Y-V-E-----A-EVS-I-----N 67
388 LDYIEVDYEGKGRQTIQ-GILIEHLYGLTFENLYATNSDANAQOKTSVIRVRFN 446
68 -E-----D-GTL-----72
447 STEYQVTRVDKGGALHYHORRQPRVRSACENDQYKPGGCDICLLANSHKARTCR 506
73 -F-----DG-----R-----PIE-----SL-81
507 RSGFSLGDSKCKPEHELFLVYKGRPGIIRGMDGAKVPDEHMPIDENLNPALDF 566
82 -----S-LI-----D-----AV-M-----PD--90
567 HAETGFIYPADTTSYILIGROKIDGTERITLKGDIHNVEGVAVDWMDNLYWTDDGPKT 626
91 V-V-----QTR-----Q-97
627 ISVARLEKAAQRTLLIEGKHTHPRAIVDPLNGMYWTWDEEDPKDSRRGRLERAMDG 686
98 -----A-Y-R-----D-K-----102
687 SHRDIFVTSKVLWPNGLSLDIPAGRLYWDAFYDRIETILLNGTDKIVYEGPELNHAF 746
103 -----L-----AQQAA-----109
747 GLCHHGYLFWTEYRSGSVYRLRGVGGAPPTVTLRSEPPPIFEIRMYDAQOQOVTNK 806
110 -----AA-----A-A-----A-RASQ-120
807 CRVNGGCSLCLATPGSROCAEAOVLADGVTCLANPSYVPPQCPQGEFACANSRC 866
121 -Q-----G-----S-----A-----K-----N-G-----E 128
867 IQERWKCDGNDCLDNDSEAPALCHQHTCPDRFKENNRNCIPNRWLCDGNDGSEDE 926
129 -N-T-A-----N-----G-----E-----E--N 136
927 SNATCSARTCPNPQFSCASGRICIPISWTCDDDDCGDRSDESASCAYPCTCPLTQFTCNN 986
137 G-----A-H-----T-----I-----141
987 GRGININWRCDNDGNDSEAGCSHSCSSTOFKCNHRCICPEHWTCDGNDGCGYSDE 1046
142 -AN--N-----HTD-----M-----E-----V-D-152
1047 THANCINQATPPGGCHTDFQCRLDGLCIPLRWRCDGDDTCDMDSSDEKSCGTHVCDP 1106
153 -G-D-----V-----E-I-----PP-----NKAV-----V-164
1107 SVRFCKDSARCISKAWVCDGNDGNDCEDNSEENCESLACRPPSHPCANNSTVCLPDKLC 1166

QY 165 -----L-R-----G-H-----E-S-----E-----V---172
Db 1167 DGNDGCGDSDEGELCQCQSLNNGGCSHNCVAPGEGIVCSGFLGMDGPDNHTCQIQSY 1226
QY 173 -----F-I-CA-----W-----N-----178
Db 1227 CAKHLKCSQKQCKFVKSCVCEGVWLEPBDGSCRSLODPFKPFIIFSNRHEIRRIDLHK 1286
QY 179 -----P-----V-----S-----D-----LL-----A-S-----186
Db 1287 GDYSLVPLGRLNTIALDFHLSQALYWTVDVEDKIYRGKLLDNGALTSFEWVIQYGLATP 1346
QY 187 -----G-----S-----GD-----S-----T--192
Db 1347 EGLAVDWIAGNIYVESNLDQIEVAKLDGTLRTLLAGDIEHPRAIALDPRDGLFTWDW 1406
QY 193 -A--RI-----W-N-----197
Db 1407 DASLPRIEASMSGAGRRTVHRETSGGWPNGLTVDYLEKRILWIDARSDAIYSARYDGS 1466
QY 198 -----LS-----E-----NST-----SG-----S-TQ--L 209
Db 1467 GHMEVLRGHEFLSHPPFAVTLYGGEVYWTDRTN-TLAKANKMTGHNVTVVQRTINTOPFDL 1525
QY 210 -V-----L-R-----H-C-----214
Db 1526 QVTHPSRQWPNPCBANGGQPCSHLCLINYNRTVSCACPHLMKHLKONTTCYEPKFL 1585
QY 215 -----I-----RE-----217
Db 1586 LYARQMEIRGVLDLAPYVNIISFTVPDIDNVTVDYDAREQVYWSVDVTRQAIKRAFIN 1645
QY 218 G-G-----QDVPSNK-----D-V-----TS-----ID-----232
Db 1646 GTGVETVWSADLP-NAHGLAVDWVSRNLFWTSYDTNKKOINVARLDGSKFNAVVQGLEQP 1704
QY 233 -----W-N-S-----EG-----TLL-----A-TG-----S-Y-----245
Db 1705 HGLVHPLRGKLYWTDGDNISMANMDGNSNRTLLFSQKQGPVGLAIDFPESKLYWISSGNH 1764
QY 246 -----D-GF-----A-R-----I-----W-----T-K-DG-----256
Db 1765 TINRCLDGSGLVIDAMRSQLGKATALAIMGDKLWADQVSEKMGTCRADGSGSVLR 1824
QY 257 N--LA-----S-----T-----L-GQ-264
Db 1825 NSTTLYMHMKVDESIOLDHKGTPNCSVNNGDCSQLCPTSETTRSCMCTAGYSLRSGQ 1884
QY 265 -----HKG-----P-----I-F-A-----L-272
Db 1885 ACEGVCSFLLYSVHEGIRGIPDPNDKSDALVPVSGTSLAVGIDFHAENDTIYVWDMGLS 1944
QY 273 -----K-W-----N-K-----KGN-----F-I-----L-----S 283
Db 1945 TISRAKDQTRBDVVVTNGIGRVEGIAVDWIAGNIYWTGQGFVIEARLNGSFYVVIS 2004
QY 284 AGVDK-----T-----T-----I-----I-----292
Db 2005 QGLDKPRAITVHEKGYLFWTEWGPYRIERSRLDGTERTVVLVNVNISWPNGISVDYQDG 2064
QY 293 -W-DA-----H-----296
Db 2065 KLYWCARDTKTERIDLETGENREVLSSNNMDMFSVPEDFIYWSDRTHANGSIKRG 2124
QY 297 -----TG-----298
Db 2125 KDNATDSVPLRTIGIVQLKDIKVFENRDRQGTNVCAVANGCCQQLCLYRGRGORACAH 2184
QY 299 -----E-A-----K-----Q-----Q-----F---304
Db 2185 GMLAEDGASCREYAGYLLYSERTILKSIIHLSERNLNAPVQPPEDPEHMKNVIALAFDYR 2244

QY 305 -----P-F-H-----S-A----- 309
Db 2245 AGTSPGTNRIFFSDIHFGNIQIINDGSRITIVENVGSVEGLAYHRGWDLTWYTYT 2304
QY 310 -----P-AL-----D-V-D-----W-Q-----S 318
Db 2305 STITRHTVQTRGAFERETVITMSGDHPRAPVLDECONLMFTWNNBOHPSIMRAALS 2364
QY 319 -N-----N-----T----- 321
Db 2365 GANVLTLIEKDIRTPNGLAIDHRAELKYPDSATLDKIERCEYDGSRYVILKSEPVHPFG 2424
QY 322 -----F-----A-----SC-S----- 326
Db 2425 LAVYGEHIFWTDVWRAVQRANKHVGSMKLLRVDIPQOPMGIIAIVANDTNSCELSPCRI 2484
QY 327 -----T-----DM-----CI----- 331
Db 2485 NNGGQDLCLLTHQHVNCSCRGRILODLDLTCRAVNSCRAQDEPECANGECINFSLTC 2544
QY 332 -H-----V-----C-K----- 335
Db 2545 DGVPHCKDKSDEKPSYCNRRCKTFRQCSNGRCVSNMLWCNAGDDCGDGSDBIPCNKTA 2604
QY 336 -----Q-----LG----- 338
Db 2605 CGVGEFRCDGTGIGNSSRCNQFVDCEDASDENMCSATDCSSYFRLGVGVLPQPCERTS 2664
QY 339 -----D-----RP-----I-----K----- 343
Db 2665 LCYAPSWVCDGANDCGDYSDERDCPGVGRPCPLNYFACPSGRCI PMSWTCKDEDDCEHG 2724
QY 344 -T-----F-----O-----G-H-----T 349
Db 2725 EDETHCNKFCSEAFQEQNHRCISKQWLCDGSDGDDGSDGDEAAHCBGKTCGPSSFSFCPGT 2784
QY 350 -----N-----EV-N-A-I-K----- 356
Db 2785 HVCVPERWLCDDKADGADESIAAGCLYNSTCDDREFMCONRQCI PKHFVCDHBRDCA 2844
QY 357 -----W-----D-----P-T----- 360
Db 2845 DGSDESPECEYPTGPFSEFRANGRLSSRQWPCDGENDCHDSEAPKNPHCTSPCHKC 2904
QY 361 -----G-N-----L----- 363
Db 2905 NASSQFLCSSGRVABALLCNGQDDCGDSDDERGCHINECLSRKLSGCQDCEDLKIGFK 2964
QY 364 -L-----A-S-CS-----D----- 369
Db 2965 CRCRPGRLKDDGRTCADVDECSTTFPCSORCINTHGSYKCLVEGYAPRGDPHPSCKAV 3024
QY 370 -D-----M-----T-LK-----I-W-----SM-- 378
Db 3025 TDEEPFLIFANRYLRKLNLDGNSYTLKQGLNNAVALDFYREOMIYWDVDTTQGSIMR 3084
QY 379 -K-----Q-----D-N-C----- 383
Db 3085 RMHLNSNVQVLHRTGLSNPDGLAVDWGVGNLYWCDKGRDTIEVSKLNGAYRTVLVSSGL 3144
QY 384 -----V-----H----- 385
Db 3145 REPRALVVDVQNGLYWTDWGDHSLGRIGMDSSRSVIVDTKITWPNGLTLDYVTERIY 3204
QY 386 -D-----L-Q-----Q-H-----N-K----- 392
Db 3205 WADAREDIIEFASLDGSRHVVLSQDIPHIFALTFLFEDYVYWDWETKSNRAHKTGTN 3264
QY 393 E-I-----Y-- 395
Db 3265 KTLISTLHRPMDLHVHALRQDPVPHPCVKNVNGCSNCLLSPGGHKACPTNFIYG 3324
QY 396 -----T-----IK-W-----S-----P-----TG 403

RESULT 2

Db 3325 SDGRTCVSNCTASQFVKNDKCIPIFWMKDTEDDCGHSDDEPPDCPEFKRPGQFCSTG 3384
QY 404 -P-----G-----TN-----N----- 408
Db 3385 ICTNPAFICDGNDCODNSDEANCDIHVCLPSQFKCTNTNRCIPGIFRCNQDNCGDGED 3444
QY 409 -----PN-----ANL-M----- 414
Db 3445 ERDCPEVTCAPNQFQCSITKRCIPRVWVCDRNDVDSDEPANCTQMTGVDDEFCKDS 3504
QY 415 -L-A----- 416
Db 3505 GRICIPARWKCDGEDDCGDSDEPKCECBRTCEPYQFRCKNNRCVPRWQCDYDNDCGDN 3564
QY 417 -----S-----A-S-----F----- 420
Db 3565 SDEESCTPRPSESESEFSCANGRCIAORWKCDGHDHDCADGSDKDCPTPCMDMOFQCKSGH 3624
QY 421 -----D-S-----T-VR-----L-W-----D 428
Db 3625 CIPLRWCDADADCMDSDEEACGTGVRTCLPDEFQCNNTLCKPLAWKCDGEDDCGNSD 3684
QY 429 -----V-----DRGIC-I-----HTLTKH- 441
Db 3685 ENPEECARFVCPNRPFRCKNDR-VCLWIGRQCDGTDCGDTDEEDCEPPTAHT-T-HC 3741
QY 442 -----Q-----E-P-----V-- 445
Db 3742 KDKKEFLCRNQRCLSSSLRCNMPFDDCGDGSDEEDCSIDPKLTSCATNASICGDEARCVRT 3801
QY 446 -Y-----S-V-----A-F----- 450
Db 3802 EXAAYACACSGFTVPGQPCQDINECLRFGTCSQLCNNTKGHLCSARNFMKTHNTCK 3861
QY 451 -----S-----P-----D-----GR----- 455
Db 3862 ABGSEYQVLYIADNNEIRSLFPGHPSHSAVEQAQFGDSVRIDAMDVHVKAGRVYTNWHT 3921
QY 456 -Y-L-A-S-----G-----S-----F-----D- 463
Db 3922 GTISYRSLPAPPTTSNRRHQIDRGVTHLNIISGLMRPGIAIDWVAGNVYWTDSGRDV 3981
QY 464 -K-C-----V-H-I-----W-N-----T-- 471
Db 3982 IEVAQMKGENRKTLSGMIDEPHAIWVDPRLGTMWSDHGNHPKIEAAMDGTLRELTVQ 4041
QY 472 -Q-TGAL-V-HS-Y-----R-GT-----G----- 484
Db 4042 DNIQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPIVAADSKRGLSHPSFIDVFED 4100
QY 485 -G-----I-F-----EV-----C-W-- 491
Db 4101 YIYGVTYINNRVFKHKFGHSPVLNLTGGLSHASDVVLVYHQHKQPEVTNPDCRKKCEWLC 4160
QY 492 -----N-----A-----A----- 494
Db 4161 LLSPSGPVCTCPNGKKLDNGTCVVPSPPTPPDAPRPGTCNLCQFPNGSGCFLNARQPKC 4220
QY 495 -----GDK-----V-G-A-- 500
Db 4221 RCOPRYTGDKCELDQCWEHCRNGGTCAASPSGMPTRCPCPTGTPKCTQOVACAGYCANNS 4280
QY 501 -----S-----ASDGS----- 506
Db 4281 TCTVNGNQPOQRCLPGFLGDRQYRQCSCYCFENFCTCQMAADGSRQCRCTAYFEGRCE 4340
QY 507 V-----CVL-----DL-----R-----K 514
Db 4341 VNKCSRLEGACVVKQSGDVTCTCNDTDRVAPASCLTCVGHCSNGSGCTWNSK 4392

US-08-469-658-52					
; Sequence 52, Application US/08469658					
; Patent No. 5917018					
; GENERAL INFORMATION:					
; APPLICANT: Th egersen, Hans Christian					
; APPLICANT: Holtet, Thor Las					
; APPLICANT: Eterodt, Michael					
; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF					
; TITLE OF INVENTION: PROTEINS					
; NUMBER OF SEQUENCES: 58					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: Fish & Richardson P.C.					
; STREET: 225 Franklin Street					
; CITY: Boston					
; STATE: Massachusetts					
; COUNTRY: USA					
; ZIP: 02110-2804					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Floppy disk					
; COMPUTER: IBM PC compatible					
; OPERATING SYSTEM: PC-DOS/MS-DOS					
; SOFTWARE: PatentIn Release #1.0, Version					
; SOFTWARE: #1.25					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/08/469,658					
; FILING DATE: June 5, 1995					
; CLASSIFICATION: 530					
; PRIOR APPLICATION DATA:					
; APPLICATION NUMBER: 08/192,060					
; FILING DATE: February 4, 1994					
; CLASSIFICATION: 530					
; ATTORNEY/AGENT INFORMATION:					
; NAME: Paul T. Clark					
; REGISTRATION NUMBER: 30,162					
; REFERENCE/DOCKET NUMBER: 06363/002002					
; TELECOMMUNICATION INFORMATION:					
; TELEPHONE: 617 542 5070					
; TELEFAX: 617 542 8906					
; TELEX: 200154					
; INFORMATION FOR SEQ ID NO: 52:					
; SEQUENCE CHARACTERISTICS:					
; LENGTH: 4544 amino acids					
; TYPE: amino acid					
; STRANDEDNESS:					
; TOPOLOGY: linear					
; MOLECULE TYPE: protein					
US-08-469-658-52					
Query Match 70.5%; Score 2580.5; DB 2; Length 4544;					
Best Local Similarity 10.1%; Pred. No. 2.4e-32;					
Matches 434; Conservative 61; Mismatches 12; Indels 3805; Gaps 385;					
QY	1 MS-----I-----SSDE-----V-----N--F-L-----11				
Db	88 MSRLCNGVDQCMDSDEGPHCEHRCQLGNC SRLCGHCVPFTLDGPTCYCNSSFQLQADGKT 147				
QY	12 -----VV-----R-----Y-LQ-----E-----18				
Db	148 KXDFFDCSVIGTCSQLCTWTDGSFGTCGVGYLLQPDNRSCAKNEPVDPRPPVLIANSQ 207				
QY	19 -----SG-----FS-----H--SA-----F 26				
Db	208 NILATYLSGAQVSTIPTSTRTQTAMDFSYANETVCWHVGDSAAQTOLKARMPGLKGF 267				
QY	27 -----TFGIK-S-HISQ-----S-N-----I-----NG-----41				
Db	268 VDEHTNISLSLHHVEQMAIDLTCGNFYFVDDIDIRIFVCNRNGDTCVTLLDLLELYNPKG 327				
QY	42 -AL-----VP-----P-----A-A-48				
Db	328 IALDPAMGKVFTDYQIPKVERCDMDGONRTKLVDSKIVPFHGITLDLVSRLYMWADAY 387				
QY	49 --LI-----SIIOKG-L-Q-Y--V-E-----A-EVS-I-----N 67				

Db 1467 GHMEVLRGHEFLSHPPFAVTLTYGGGVYWDWRTN-TLAKANKTGHNVTVVQRTNTPQFDL 1525
Qy 210 -V-----L---R-----H-----C----- 214
Db 1526 QVYHPSRQWPNPCEANGQGPCSHLCLINYNRTVSCACPHLMKHLKONTTCYEPKFKL 1585
Qy 215 -----I-----RE----- 217
Db 1586 LYARQMEIRGVLDAPYNYIISFTVPDIDNVTLVDYDAREQRYWMSDVRTQAIKRAFIN 1645
Qy 218 G-G-----ODVPSNK-----D-V-----TS-----LD----- 232
Db 1646 GTGVETVVSADLP-NAHGLAVDWVRNLFWTSYDXTKKQINVARLDGSGFNNAVVGLEQP 1704
Qy 233 -----W-----N-S-----EG-----TL-----A-----TG-----S-Y----- 245
Db 1705 HGLVHPLRCKLYWTDGDNISMANMDSNRTLLFSGQKGPVGLAIDPPESKLYWISGNH 1764
Qy 246 -----D-GF-----A-R-----I-----W-----T-K-DG----- 256
Db 1765 TINRNLGSGLEVIDAMRSQLGKATALAIMGDKLWADQVSEKMGTCCKADGSGSVLR 1824
Qy 257 N---LA-----:-----:-----S-----T-----L---GQ- 264
Db 1825 NSTTLMHMKVYDESQLDHKGTPNCSVNNGDCSQLPTSETTRSCMCTAGYSLRSGQ 1884
Qy 265 -----HKG-----P-----I-F-A-----L- 272
Db 1885 ACEGVSGFLLYSHEGIRGIPLPNDKSDALVPVSGTSLAVGIDFPAENDTIYVDMGLS 1944
Qy 273 -----K-----W-----N-----K-----KGN-----F-I-----L-----S 283
Db 1945 TISRAKQDQWREDVVTNGIRVEGIAVDIAGNIYWDQGFVIEVARLNGSFRYVVIS 2004
Qy 284 AGYDK-----T-----T-----I-----I----- 292
Db 2005 QGLDKPRAITVHPEKGYLFTWEGQYPIERSRLDGTERTVYLVNVSISPNGISVDYQDG 2064
Qy 293 -----W-DA-----:-----:-----H----- 296
Db 2065 KLYWCDAKTDKIBRIDLETGENREVLSNNMDMFSVPEDFIYWSDRTHANGSIKRG 2124
Qy 297 -----TG-----:-----:----- 298
Db 2125 KDNATDSVPLRTGIGVQLKDIKVPNRDQKGTNVCAVANGCQCLYRGORACACAH 2184
Qy 299 -----E-A-----K-----Q-----Q-----F----- 304
Db 2185 GMLAEDGASCREYAGYLLYSERTILKSIHLSDERNLNAVPQPPEDPEHMKNVIALAFDYR 2244
Qy 305 -----P-----E-H-----H-----S-----A----- 309
Db 2245 AGTSPGTNRIFFSDIHFGNIQIINDGSRRIIVENVSGVEGLAHRGWDLYWTSYTT 2304
Qy 310 -----P-AL-----D-----V-D-----W-Q-----S 318
Db 2305 STITRHTVQTRGAFERETVITMSGDDHPAPFVLDECNLMFTWNEQHPSIMRAALS 2364
Qy 319 --N-----N-----T----- 321
Db 2365 GANVLTLIEKDIRTPNGLAIDHRAEKLIFYSDATLDKIERCEYDGSYRILKSEPVHPFG 2424
Qy 322 -----P-----:-----:-----A-----SC-S----- 326
Db 2425 LAVYGEHIFWTDWVRRAVQRANKHVSNNKLLRVDIPQPMGIIVANDTNSCELSPCRI 2484
Qy 327 -----T-----DM-----CI----- 331
Db 2485 NNGGQDCLCLTHQGHVNCRCGRILQDDLTCTRAVNSSCRAQDEFCANGCEINFSLTC 2544
Qy 332 -----H-----:-----:-----V-----C-K-- 335
Db 2545 DGVPCHKDKSDEKPSYCNRRCKKTFRCQSNRCVSNMLWCNGADDCGSDGSDBIPCNKTA 2604

Qy 336 -----LG-----Q----- 338
Db 2605 CGVGEPRCRDGTGIGNSSRCNQFVDCEDASDENMNCSATDCSSYFRLGVKGVLPQPCRTS 2664
Qy 339 -----D-----RP-----I-----K----- 343
Db 2665 LCYAPSVWCDGANDCGYDSRDCPVGKPRCPPLNYFACPSGRICIPMSWTCDKEDDCEHG 2724
Qy 344 -----T-----F-----Q-----G-----H-----P-T----- 349
Db 2725 EDETHCNKFCSEAFQECQNHRCISKOWLCGSDGSDGDEAAHCEGKTCGPSFSFCPGT 2784
Qy 350 -----N-----EV-----N-A-I-K----- 356
Db 2785 HVCVPERWLCDDKDCADGADESIAAGCLYNSTCDDREFMCMQNRQCIPIKHfVCDHDDCA 2844
Qy 357 -----W-----D-----P-T----- 360
Db 2845 DGSDESPECEYPTCGPSEFRANCGRCLSSRQWECGENDCHQSDBAKPNPHCTSPHKC 2904
Qy 361 -----G-----N-----L----- 363
Db 2905 NASSQFLCSSGRCAEALLCNGODDCGSDSDERGCHINECLSKLSGCSQDCBDLKIGFK 2964
Qy 364 -----L-----A-----S-----CS-----D----- 369
Db 2965 CRCRPGFRLKDDGRTCADVDDECSTTFPCSQRCINTHGSYKCLCEGYAPRGGDPHSCAV 3024
Qy 370 -D-----M-----T-LK-----I-W-----SM-- 378
Db 3025 TDEEPPLIFANRYLRKLNLDGSNTLLKQGLNNAVALDFDYREQMIYMTDVTQGSIMR 3084
Qy 379 -K-----Q-----D-----N--C----- 383
Db 3085 RMHLNSNVQVLRHTGLSNPDGLAVDWGGLNYCWDKGRDTIEVSKLNGAYRTVLVSSGL 3144
Qy 384 -----V-----H----- 385
Db 3145 REPRALVVDVQNGLYWTDWGDHSLIGMDGSSRSVIVDTKITWPNGLTLDYVTERIY 3204
Qy 386 -----D-----L-----Q-----H-----N-----K----- 392
Db 3205 WADAREDTIEFASLDGSRHVVLSQDIPHFALTLPEDYVYWTDMETKSNRAHKTGTN 3264
Qy 393 E--I-----:-----:-----Y-- 395
Db 3265 KTLTLLSTLHRPMDLHVHFRALRQDPVNHPCVNNNGCSNCLLSPGGHKCACTPNFYLG 3324
Qy 396 -----T-----IK-W-----S-----P-----TG 403
Db 3325 SDGRTCVSNCTASQFVCKNDKCLIPFWWKCDTEDDCGDHSDPEPDCPFKCRPQFQCSGTG 3384
Qy 404 -----P-----G-----TN-----N----- 408
Db 3385 ICTNPAPICDGNDCQDNDSEANCDIHVCLPSQFKCTNTRCIPGIFRCNGQDNCGDGED 3444
Qy 409 -----PN-----ANL--M----- 414
Db 3445 ERDCPEVTCAPNQFQCSITKRCIPRVWVCDRDNDCVDGSDPANCTQMTGCVDEFRCQDS 3504
Qy 415 --L-A-----:-----:----- 416
Db 3505 GRICIPARWKCDGEDDCGSDGSDPEKCECBERTCEPYQFRCKNNRCVPRWQCDYDNDGDN 3564
Qy 417 -----S-----A--S-----F----- 420
Db 3565 SDEESCTPRPSESEFSEFCANGRCIAGRWKCDGHDGDCADGSDKDCDTPRCMDQFQCKSGH 3624
Qy 421 -----D-S-----T-VR-----L-W-----D 428
Db 3625 CIPLRWRCADADCMGSDGDEEACGTGVRTCPLDFEQCNNTLCKPLAWKCDGEDDCGSDNSD 3684

QY 429 -----V-----DRGIC-I-----HTLTKH- 441
Db 3685 ENPECAFCVCPNRPCKNDR-VCLWIGROCDGTNCGDGTDEEDCEBPTAHT-T-HC 3741
QY 442 -----Q-----E-----P-----V-- 445
Db 3742 KDKKEFLCNRQCLSSSLRCNMFDDCGDSEEDCSIDPKLTSCATNASICGBEARCVRT 3801
QY 446 ---Y---S---V-----A-F----- 450
Db 3802 EKAAYACACRSFGHTVPGQPGCODINECLRFCTCQLCNNTKGHLCSARNFMKTHNTCK 3861
QY 451 -----S-----P-----D-----GR----- 455
Db 3862 AEGSEYOVLYIADNEIRSLFPGHPSAYEQAFQGSDESVRIDAMDVHVKAGRVYTNWHT 3921
QY 456 ---Y-L-A-S-----G-----S-----F-----D- 463
Db 3922 GTISYRSLPPAAPPTTSNRHRRQIDRGVTHLNISGLKMPRGIALDWAGNVYVTDSGROV 3981
QY 464 ---K-C-----V--H-I-----W-N-----T--- 471
Db 3982 IEVAQMKGENRKTLLISGMIDEPHAIIVVDPLRGTMWSDGNHPKRIETAAMDGTLRETLVQ 4041
QY 472 ---Q-TGAL-V-HS---Y-----R-GT-----G----- 484
Db 4042 DNTQWPTG-LADVHNERLYWADAKLSVIGSIRLINGTDPPIVAADSKKGLSHPSIDVFED 4100
QY 485 ---G-----I-F-----N-----A----- 494
Db 4101 YIYGVTYINNRFKIHKFGHSPVNLTLGSLHASDVVLYHGHQKPEVTNCDRKKCEWLC 4160
QY 492 -----N-----A----- 494
Db 4161 LLSPSGVCPCPNKRLDNGTCVVPSPPTPPDAPREGTCNLCQFNGSGCFLNARRQPKC 4220
QY 495 -----GDK-----V-G-A--- 500
Db 4221 RCPRYTGDKCELDQWEHCNNGTCAASPSGMPTRCPTGPTGPKCTQVCAGYCANNIS 4280
QY 501 -----S-----ASDGS----- 506
Db 4281 TCTVNOGNOQCRLPGFLGDRCOYRCQSGYCNFGTCOMADGSRQCRCTAVFEGRSCE 4340
QY 507 V-----CVL-----DL-----R-----K 514
Db 4341 VNKSRCLGACVYVVKOSGVDTCNCTDGRVAPSLCTCVGHCSNGSGCTMNSK 4392

RESULT 3

US-09-334-220-2
; Sequence 2, Application US/09334220
; Patent No. 6323177
; GENERAL INFORMATION:
; APPLICANT: St. Jude's Children's Research Hospital
; APPLICANT: Curran, Thomas
; APPLICANT: D'Arcangelo, Gabriella
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND
; TITLE OF INVENTION: THERAPIES
; FILE REFERENCE: 2427/0F704
; CURRENT APPLICATION NUMBER: US/09/334,220
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 3461
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-334-220-2

Query Match 70.3%; Score 2572.8; DB 3; Length 3461;
Best Local Similarity 11.8%; Pred. No. 7.7e-33;

Matches 395; Conservative 90; Mismatches 16; Indels 2842; Gaps 361;
QY 1 M-----S-----S-----S-----D-----EV--- 8
Db 1 MERGCWAPRALVLAVILLATLARAATGYYPFPFFLCTHHGBLEGDEGEVLISL 60
QY 9 ---N-----F-LV--YRY--LQ--ES-----GF--S-H 23
Db 61 HIAGNTYVYVGOEYHVTISTSTFFDGLLVTL-YTSTSIQSSQSIGSSAFGFGMSDH 119
QY 24 -----S-A-----F-----T-----G-I-K----- 31
Db 120 QFGNQFMCWSWASHVSHLPTTNLSFWVIAPAGTGCNFMATATHRGOVIFKDALAQOLC 179
QY 32 -----SH-----I-----S-Q-----SNI-----N-----GAL----- 43
Db 180 EQGAPTEATAYSHLABIHSDSVILRDDFDYSYOQLNPNINWECSCNCEMGEQCGTTHMGN 239
QY 44 -V-----P-P-----A-AL--I-----SI-----I-Q-- 54
Db 240 AVTFCBPYGPRLTTTCLNTTASVLQFSIGSGSCRFSDPSITVSYAKNNTADMIQLE 299
QY 55 -----KG-----LQY-----V-EA-E-----VSIN----- 67
Db 300 KIRAPSNVSTVIHILYLPBEEKESVQFQWKQDSLVRGVEYACWALDNILV-INGAHRE 358
QY 68 ---ED-----GT-IF-----DG-----R----- 76
Db 359 VVLEDNLDPDVTGNWLFPPGATVKHSCQSDGNSIYFHNESGSEFNATTTRDVLSTEDIQ 418
QY 77 -----P-----I-----ES-LSLI-----DA----- 86
Db 419 EQWSEBFEQPTGWDILGAVVGADCTVESGSLVFLKDGKRLCTPYMDTTCYGNLRFY 478
QY 87 -VM-----P-----D-----D-----VV-----QT--- 94
Db 479 FVWGIGCDGVSHENDIILVAKIEGRKEHIALDTLTYSSYKVPVSLVSVINPELQTPATK 538
QY 95 ---RQ---QAY-R---D-----K-LAQ-----Q----- 106
Db 539 FCLROKSHOGYNNRWAVDFHVLVPLPSTMHMIQFSLNLCGTHQPGNSVSLFSTNH 598
QY 107 -QA-----A-----AAA-----AA----- 114
Db 599 GRSWSLLHTECLPEICAGPHLPSTVYSSNYSYGNWNRITPLPNAALTRDIRWRQTGP 658
QY 115 -----A-----A-----A-----A-----A-----Q--GS 123
Db 659 ILGNWMAIDNVYIGPSCLKFCSGRGQCTRHGCKCDPFGSPACEMASQTFPMFISEBFGS 718
QY 124 AK-----N-----G-E-----N-----G- 129
Db 719 ARLSSYHNFYSIRGAESVFCGVLASGKALVFNKGRROLITSLDSSQSRFLQTLRLG 778
QY 130 -----T-A-----NG--E-E--E-----N-----G- 137
Db 779 SKSVLSTCRAPDQPGEGVLLHYSDNGITWKLLEHYSYVNYHEPRIISVELPDDARQFGI 838
QY 138 -----A-H-----T-----I-----AN-----N-----H- 145
Db 839 QFRWQPYHSSQGEDVWAIDEIVMTSVLFNSISLDFNLVEVTQSLGFLVGNVQPYCGHD 898
QY 146 -----T-D-----M-----NEVDG-----D--V--E----- 156
Db 899 WTLCTFTGDSKLASSMRVFTQSMQI-GASYMTQFSLVMGCGQKYTPHMDNQVKLEYSANH 957
QY 157 -----I-----P--EN--K-----A-----VVL-----R----- 166
Db 958 GLTWHLVQBECLDPMFSCQEFTSASIHASEFTQWRRTVVVLPOKTWSGATRRWSQSY 1017
QY 167 -----GH-----E-----E-SE-----V---F--- 173
Db 1018 TAQDEWALDNIVYIQCCPNMCSGHGSCDHGVCRCDOGYQGTCHPEAALPSTIMSDPENP 1077

QY 174 -----I-----C-----A-----W----- 177
Db 1078 SSWESDQWQVIGGEWKEQCGVSSGSLYFSKAGKQQLVSWDLDTSVDFVQFYQI 1137
QY 178 -----N-P-----V-----SD-----L-L-ASG----- 187
Db 1138 GGSAACNKPDSREBEGILLOYSNNGIQWHLAEMFYDFSKPRFVYVLELPAAGKTPCTR 1197
QY 188 -----SG-----D-----S-----T-----A-----R-IW- 196
Db 1198 FRWKPVSCEDDQWAVDDIILSEKQKQVIPPVNPPLPQNFYKPAFDYPMNQMSVWL 1257
QY 197 ---N--LSEN-S---T-S---G-S---T-----Q-L----- 209
Db 1258 MLANEGWAKNDSFCATTPSAMVFGKSDGRFAVTRDLTLKPGVYLQKLNIGCTSQFSST 1317
QY 210 --V-LR--H-----C-----IR-EG-----GQD-----V 222
Db 1318 APVLLQYSHDAGMSWFLKKEGCFPASAAKCEGNSRELSPTVYTYG-DPEEWTRITIAI 1376
QY 223 P-----S-N-----K-----D----- 227
Db 1377 PRSLASKTRFRWIOESSQKNVPPFGLDGYIISEPCPSYCSGHGDCISGVCFDLGYTA 1436
QY 228 ---V--T--S--LD-----W-----NSEG-TL-----L--A-- 241
Db 1437 AQGTCVSNTPNHSEMPDRFEGKSLPLWKYKITGGQVGTGCGTLN-DGRSLYFNLGKREAR 1495
QY 242 T-----GS-----Y--D-G-----F 248
Db 1496 TVPLDTRNISLVQFYIQIGKSGITTYTPRAREYGLVQVSNNDGILMHLLELDFMSF 1555
QY 249 -----A-RW-----T-----K-DG- 256
Db 1556 LEPOIISIDLPREAKTATAFR-WWOPQHKHSAQWALGVILGVNDSQTGFQDKLDGS 1614
QY 257 -----N-----LAS--T-L--G--Q-----H-----KG- 267
Db 1615 IDLQANWYRIQGOVDIDLCL-SMDTALIFTENIGNPRYAEWTDPHVSSESLQWENMGC 1673
QY 268 --PIP-A--L-----K-W--NKK-----G----- 278
Db 1674 SKP-FSGAHGIQLOYSLNKGQWLTVECPPTIGCVHYTESSTVTSERFONWRRVTYV 1732
QY 279 -----N-----F--I--LSAGYDKT-TI--I-----W-----DA----- 295
Db 1733 LPLATNSPRTRFRWIOQNTYTVGAD-SWAIDNVILASGCPMWCGRGICDGRVCDRFGF 1791
QY 296 -----H-----T-----CE----- 299
Db 1792 GPFCVPVPLPSILKDDPFGNLHPDLWPVYVGAERGNLNGETIKSGTCLIFKGEGRMLI 1851
QY 300 -----AK-Q--QF-----P-----F-----H-----S----- 308
Db 1852 SRDLCTNTWYQVSLRFTAQGTPEKSHSILLQFSVSGVTVHLMDBEFPQPTTSLFIN 1911
QY 309 AP-----A--L-----D-----V-----D-W 316
Db 1912 VPLPYGAQTNATFRMLQWPNYNGKBEIWIIDDFIIDGNLNNPVLILLDTDFGPREDNW 1971
QY 317 -----Q--S-N-NT----- 321
Db 1972 FFYPGGNIGLYCPSYSGKAGEEDSAMVFSVNEVEHSEITTRDLNVNENTIIQPEINVGCS 2031
QY 322 -----F-A-----S-CST----- 327
Db 2032 TDSSADPVRLEFRSDFGATWHLPLCYHSSSLVSLSCSTEHPHSSTVYAGTTQWRRE 2091
QY 328 -----D-----MC-IH-----VC 334
Db 2092 VWHFGKLHLCGSVRFRWYQFYAGSPQVMTWIDNVYIGQCCEMICYGHGSCINGTKCIC 2151

QY 335 -----K--L-----GO--DR-----P-----I-KT-----P-Q-G- 347
Db 2152 DPGYSGPTCKISTKNPDFLKDDPEQGLSDRFLMLSGGKPSKRCGILSSGNLFFNEDGL 2211
QY 348 -----H-----T-N-----EV----- 352
Db 2212 RMLVTRDLDLSHARFVQFPMRLCGKGVDPDRSQPVLLQYSLNGLSWSLQBLFNSNS 2271
QY 353 NA--I--K-----W----- 357
Db 2272 NVGRYIALEMPKARSSTRLRWQPSENGHFYSPWVIDOILIGGNISGNTVLEDDFSTL 2331
QY 358 D-----P--TG-----N--LL-----ASCS--D 369
Db 2332 DSRKLLHPGGTRMPVCGSTGDALVFEKASTRYVVTDDIAVNEDSLQIDFAASCSTVD 2391
QY 370 -----DM-----T-----L-KI-----W----- 376
Db 2392 SCVAIELEYSVDLGLSMHPLVRDCLPTNVECSRYHLQIRILVSDTFNKTWTRITLPLPSYTR 2451
QY 377 S-----M-----KO-----DN-----CV-----H-----D-- 386
Db 2452 SQATRRWHQAPAPFDKQOTWADNVYIGDCLDMCSGHRCVCGSCVDEQWGLYCDPE 2511
QY 387 --L--Q-----Q-----H----- 390
Db 2512 ETSLPTQLKDNFRAPSNQNLTVSGKLSLTVCGAVASGLALHFGSGCSRLLVTDNLNT 2571
QY 391 NKE-I--Y--T-----IKW-----S-PTG-----P 404
Db 2572 NAEFIQFYMYGCLITPFSNRNQVLLEYSVNGGITWLLMEIFYDQYSKP-GEVNTILLPP 2630
QY 405 ---CT---NNP-----N--A--NLMLA-SA-----SF-----D 421
Db 2631 DAKBIATRRWQO-PRHGDGLQNDWADNVILISGSADQRTVMDLTSSAPVPOHERSPAD 2689
QY 422 S-TV-R-----L-----W--DVD--R-----GI--C-IH--TL--TKH- 441
Db 2690 AGPVGRIAFEMFLEDKTYSVENWLFHD-DCTVERPCDSPDGMVLCGSHDGRVYAVT-HD 2747
QY 442 ---Q-----E-P-----V-YSAF-----S--P-D----- 453
Db 2748 LPTENWIMQFKISVGCKVPEKIAQNIHVQFSTDFGVSWSYLVPOCLPADPKCSGSVQ 2807
QY 454 -----G-R--Y-L-AS--G-----S-----F-----D- 463
Db 2808 PSVFPFTEGKRITYPLPESLTGNVFRFRPYQKYSVQWADNVYLGPGCLDNCGGHGC 2867
QY 464 -K-CV-----H-----I--W-----NT- 471
Db 2868 LKEQCICDPYSGPNCYLTHSLKTLKRFDSBEIKPDLWMSLEGGSTCTCECVLAENTA 2927
QY 472 ---Q--T-----GA--L-----V-----H-----S----- 479
Db 2928 LYFGSTVQAITQDLDLRGAKFLOYWGRIGSENNTSCHRVPCKEGLVLLDFSTDGGIT 2987
QY 480 -----Y--R-----G-----T----- 483
Db 2988 WTLLEHMDFOKYISVRHDYILLPEGALTNTTLRWQPPFVINSGLVSVGVERAQWALDNI 3047
QY 484 --GG--I-----F-----E-----V--C-----W-----NA 493
Db 3048 LIGBAEINSQLVDTFDESGSSHEENWSFYVNAVRTAGFCGNPSFHYLWPNKKDKTHNA 3107
QY 494 -----A--GD-----K----- 497
Db 3108 LSSRELIQGYMOPKIVVGCEATSGDLHSVMLEYTKDARSWSQLVQTCLPSSNS 3167
QY 498 VG-----A-----SA-----S--D-- 504
Db 3168 ICGSPQFHEATYNAVNSSWKRTITQLPDHVSSATOFRWIQKEETEKSQWADHVY 3227
QY 505 -G-----S-----VC----- 508

Db 1140 EKONSTETCOPQFQNCNHRNCRIDLSFVCDGDKDVCVDSDEVGCVLNCNTASQFKCASGDK 1199
Qy 163 ---V---VL---R-G-H---E--- 169
Db 1200 CIGVTRCDGVFDCSDNSDEAGCPTRPNGCHSDEFCQEDGICIPNFWECDGHPDCLYG 1259
Qy 170 S-E---V---P---TC---AW---N---P--- 179
Db 1260 SDEHNACVPKTCPSYFHCNCGNCRHAWLDCDRNDCGMSDEKDCPTQPFRCFSQWQC 1319
Qy 180 ---V---V---SD--- 182
Db 1320 LGHNICVNSVCDGIFDCPNGTDESPLCNGNSCDFNGGCTHCVOEPFGAKCLCPGLF 1379
Qy 183 LLA---S---GS---G--- 189
Db 1380 LLANDSKTCEDIDECILGSCSHQCYNMRGSPRSCDTGYMLESDBRTCKVTASESLLLL 1439
Qy 190 ---DS-T---A---RI---W---N--- 197
Db 1440 VASONKIIADSVTSQVHNIYSLVENSYIIVADPDFDSISGRIFWSDATQGTWAFQNGTD 1499
Qy 198 ---LSE---N---S---T---S---GS---T---Q-LVL 211
Db 1500 RRVFDSIILTETIADWVGRNLYWTDYALETIEVSKIDGSHRTVLISKULTNPRGLAL 1559
Qy 212 ---R---H--- 213
Db 1560 DPRNEHLLFWSDWGHHPRIERASMDGSMRTVIQDKIFWPCGLTIDYPNRLLYFMDSYL 1619
Qy 214 ---C---IR---E---GG--- 219
Db 1620 DYMFCDYNGHRRQVITASDLIIRHPYALTFEDSVVTTDRATRVMRANKWGGQSVV 1679
Qy 220 ---Q--- 220
Db 1680 MYNIQPLGIVAVHPKQPNVNCFAFRCRSHLCLLSQGPFPYFCVCPGSGNSLSPDLLN 1739
Qy 221 ---D---V---P---SN-K---DV--- 228
Db 1740 CLRDDQFLITVQHOIIFIGISLNPVKSDNDAWPIAGIQLGLDVEFDFAEQIYIYVWENPG 1799
Qy 229 ---T---S---LDW---N---S-E---G---T 238
Db 1800 EIHVRKTDGNTRVFASISVMGFSMLALDWISERNLYSTNPRTQSIEVLTLHGDIYRKT 1859
Qy 239 LLA---T---G---S---Y---DG---F--- 248
Db 1860 LIANDGTALGVGFPFIGITVDPARGLYWSQDGTDSGVPAKIASANMDGTSVKTLFTGNLE 1919
Qy 249 ---A---R-I---W---T--- 253
Db 1920 HLECVTLIDIEQKLYWAVTGRGVIERGNVDGTRMILVHQLSHPWGIAVHDSFLYYTDEQ 1979
Qy 254 ---K-D---G-N---L---ASTL-G-Q-H---K--- 266
Db 1980 YEVIERVDAKATGANKIVLRDNVNLQVYHRRNAESSGNSNNMACQQICLPVPGG 2039
Qy 267 ---G---P---I---FALK--- 273
Db 2040 LFSCACATGKLPNDRSCSPYNSFIWVMSLAIRGFSLELSDHSETMVPVAGQGRNALH 2099
Qy 274 ---W---N---K---K---GN- 279
Db 2100 VDVVSSGFIWYCDFSSVASDRAIRIKPDGSSLMNIVTHGIGENVRGIAVDWVAGNL 2159
Qy 280 ---I--- 281
Db 2160 YFTNAFVSETLIEVLINTYRVLKVTVMPRHIVVDPKNRYLFWADYQGPPIERSF 2219
Qy 282 ---L-S---A---G---VD--- 287
Db 2220 LDCTNRTVLVSEGIPTVPRGLAVDRSDGVVYVWVDDSLDIIRINGENSEVIRGSRYP 2279

Qy 288 ---K---T---T-I---I-W---D--- 294
Db 2280 PYGITVPENSIIWVDRNLKKIPOASKEPENTTEPTVIRDNINWLRDVTIIFDKQVQPRSPA 2339
Qy 295 ---A---HT---G---E---A--- 300
Db 2340 EVNNPCLENNGCGSHLCPALPGLHTPKDCAFGLTQSDGKNCAISTENFLIFALSNSLR 2399
Qy 301 ---K---Q---Q---F--- 305
Db 2400 SLHLDPENHSPPPQTINVERTVMSLDYSVSDRIYFTQNLASGVGQISYATLSGHTPT 2459
Qy 306 ---F--- 306
Db 2460 VIASGIGTAGIADFWITRIIYSDYLNQMINSMAGDSNRTVIARVPKRAIVLDPDQCG 2519
Qy 307 ---H---S-AP-ALD- 313
Db 2520 YLYWADWDTHAKIERATLGNFRPVIWSSLVMPGSLTLDYEEDLLYVWDASLQRIERST 2579
Qy 314 ---VD---W---Q---S---N--- 319
Db 2580 LTGVDRVIVNAVAHAFGLTLYQYIYWTDLTYQRIYRANKYDGSQOIAMTTLNLLSQPRG 2639
Qy 320 ---NT---F--- 322
Db 2640 INTVKNQKQCCNPNCEQFNGGCSHICAPNGABEQCPHEGNWYLANNRKHCIVDNGER 2699
Qy 323 ---AS---CS---TD-M---C-IH---V--- 333
Db 2700 CGASSPTCSNGRCISBEWKCDNDGCGSDMESVCALHTCSPTAFTCANGRCVQVSYR 2759
Qy 334 CKL---G---OD---R-P-I--- 342
Db 2760 CDYNDGCGSDEAGCLFRDCNATTFMCNRRRCIPREFICNGVDNCHDNTSDEKNCPD 2819
Qy 343 ---K---K---PRVYLCDGNDGDSNDENPTYCTHTCSSEFQASGRICIPQ 346
Db 2820 RYCQSGYTKCHNSNICIPRVYLCDGNDGDSNDENPTYCTHTCSSEFQASGRICIPQ 2879
Qy 347 ---GH---T--- 349
Db 2880 HWYCDQETDCFASDEBPASCGHSERTCLADEFKDCGRCIPSEWICDGDNDGDMDSDEK 2939
Qy 350 ---N---E---VNA---IK-W--- 357
Db 2940 RHQCQNQCSDEFLCVNDRPPDRRCIPQSWVCDGSDVCTDGYDENQNCNTRTCSENEFT 2999
Qy 358 ---LAS---CS---D-D---M--- 357
Db 3000 CGYGLCIPKIFRCDRHDCGYSDERGCILYQTCQOQNFQTCQNGRCISKTFVCDENDDCGD 3059
Qy 358 ---D---PT---G---N---L--- 363
Db 3060 GSDELMLHLCHTPEPTPPHEFKCDNGRCIEMMKLCHLDDCLDNDSEKKGCGINECHDPSI 3119
Qy 364 ---LAS---CS---D-D---M--- 371
Db 3120 SGCNHNCTDLTSFYGSCRPYKLMDSKRTCDVIDECTEMPFVCSQKCNENVISYICKCA 3179
Qy 372 ---T---L--- 374
Db 3180 PGLREPDGKTCRQNSNIEPYLIFSNRYLYRLNLTIDGYFSLILEGLDNVVALDFORVEK 3239
Qy 375 ---I---W---S--- 377
Db 3240 RLYWIDTORQVIERMPLNKTNETIINHRLPAAESLAVDWVSRKLYWLDARLDGLFVSDL 3299
Qy 378 ---M-KQ---D-N-C---VHDLOO---H---NK- 392
Db 3300 NGGHRMLAQHCVDANNTFCFONPRGLALH---POGYLYWADWGHRAIYGRVGMGTNKS 3357

```
QY 393 -----E-----I-YT-----IK----- 398
Db 3358 VIISTKLEWNGITIDYTNLLLYADAHGLYIEYSDLEGHRRHTVYDGPALPHFAITIFE 3417
QY 399 -----W-----S-----P-T---G-- 403
Db 3418 DTIWTDNWTRTVEKNGKDGSRQTLVNTTHRPFDIHVYPRQPIVSNPCGTNGGCS 3477
QY 404 -----PG-----T-----NN----- 408
Db 3478 HLCILKPGKGFTCECPDPDFRTLLQSGSTYCMPCMSSTQFLCANNKRCIPIMWKCDQKD 3537
QY 409 -----P-----NA-N-----LML-----AS 417
Db 3538 CSDGSDLALCPQFRLGQFCSDGNCTSPQLCAHQNCPOGSDDEDL-LCENHHDS 3596
QY 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-DV 429
Db 3597 NEWQANKRCIPESWQCDTFNDCEDNDESDSHCASRTCPGQFRCAANGRCIPQAWKCDV 3656
QY 430 --DRG-----I-CI--H-----T-L----- 438
Db 3657 DNDGDSHSDPEIECMSAHLCDNFTEFSCKTYRCIPKWAVCNGVDDCDRNSDEQCEE 3716
QY 439 -T-----K-H-----Q-----E--P-----V----- 445
Db 3717 RTCHPVGDPRCKNHCIPLRWQCGQNDGSDNDENCAPRECTESEFRVCVNOQCIPSRW 3776
QY 446 -----Y-----S-----V-----A-----F 450
Db 3777 ICDHYNCGDSDRDCMERTCHPEYFQCTSGHCVSHSELKDGSDADCLDSEADCFTRF 3836
QY 451 SPDGRYL-AS-----GS-----P----- 462
Db 3837 -PDGAYCOATWFECKNHVCIPPYWKCDGDDCGDSDDEHLCLDVECNPNFRCDNNR 3895
QY 463 -----D-----K-----CV-H--I----- 468
Db 3896 CIYSHEVCNGVDDCGDGTDETEHCKRPTKPTCYEYKCGNGHCIPHNDVCDADDDCGD 3955
QY 469 W-----N-----T-----Q-T-----G-----A-----L-V----- 477
Db 3956 WSDGLGCKGKERTCAENICQNQLNEGFFICSCTAGFTNVFDRDTSCLDINECEQFG 4015
QY 478 ---H-----SY-----R----- 481
Db 4016 TCPQHCRNKGSEYECVADGFTSMRDRPKRCAEAGSSPLLLLPDNDVIRKYNLSSEF 4075
QY 482 -----G-----T-----G-----G-----I-FE-----V-C-- 490
Db 4076 EYLODEEYIQANDYDWDPKDIGLSVYVYTVRGESRFGAIAKRAYIPNFESGRNLLVQEV 4135
QY 491 -----W-----N-----AA----- 494
Db 4136 LKLKYNMQPDGIAVDWVGRHRYNSDVKNKRIEVAKLDRYKWLITDLDQPAIAVNP 4195
QY 495 -----G-----D-----G-----D-K-V--- 498
Db 4196 LGLMFWTDWKGKPIESAMWNGSDRNILVFDLGMPTGLSIDYLNDRYWSDFKEDIET 4255
QY 499 -----G-----A-SA---S-----D-----G----- 505
Db 4256 IKYDGTDRRVIKAEAMNPYSIDIFEDQLYWISKEGVEWKNKFGQKKEKTLVVPNPLT 4315
QY 506 -----SV-----C-V-----L-----D----- 511
Db 4316 QVRIFHOLRYNKSVPNLCKOICSHLCILRLPGYSACAPQGSFTEGSTRTECDAAIELPIN 4375
QY 512 L-----R-----K 514
Db 4376 LPPPCRCMHGGNCYFDETDLPK 4397
```

```
RESULT 5
US-08-652-877-84
; Sequence 84, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Rask, Lars
; APPLICANT: Crumley, Gregg R.
; APPLICANT: Morse, Clarence C.
; APPLICANT: Murray, Edward M.
; APPLICANT: Hjaln, Goran
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
; TITLE OF INVENTION: Thereof and DNA Encoding Same
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd., 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426-0107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.5.1
; SOFTWARE: Word 6.0 (Patentin)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,877
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: PCT/US95/15203
; APPLICATION NUMBER: PCT/US95/15203
; FILING DATE: 22-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,836
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,314
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitzky, Martin
; REGISTRATION NUMBER: 29,699
; REFERENCE/DOCKET NUMBER: A1355E-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-454-3816
; TELEFAX: 610-454-3808
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4655 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-652-877-84
```

Query Match 70.1%; Score 2566.6; DB 3; Length 4655;
Best Local Similarity 9.8%; Pred. No. 5.1e-32;
Matches 433; Conservative 57; Mismatches 19; Indels 3894; Gaps 381;

```
QY 1 M-----S-----I-----S---SDE- 7
Db 1 MDRGPAVACTLLALVACLAPASGQECDSAHPRCGSHGICPADWRCDGTGKDCSDDADEI 60
QY 8 ---VN---F-----L---V----- 12
Db 61 GCAVWTCQOQYFKQSEGCIPSSWVCDQDQDCDSDRQDCSQSTCSSHQITCSNGQC 120
QY 13 ---YR---Y-----L-----Q-----E---S 19
Db 121 IPSEYKCDHVRDCPDGADENDCOYPTCEQLTCDNAGACYNYSKCDWKVCDRSDSDEINCT 180
QY 20 ---G-FS-----H-----S---A--F-T---F---G-I--- 30
```


Db 181 EICLHNEFCNGECIPRAYVCHDNDQDGSDEHACNYPTCGYQFTCPSGRCIYQNWV 240
QY 31 K-----S-----H-----IS-----O-SNI 39
Db 241 CDGEDDKDNGDEDCESGPHDVHKCSPREWSPESGRCISIKYKVCIDILDCRGREDEN- 299
QY 40 N---G---AL---VP---PAALI---S-----I-I-- 53
Db 300 NTSKGYSMTLCSALNCQYCHETPYGGACFPPIIINHNSRTEVEFDQCIWIGICD 359
QY 54 QK-----G-L---QY-----VE 61
Db 360 QKCESRPGRLHCEGEYILEROYCKANDSFGESAIIFNSGRDLLGDIHGRSFRILVE 419
QY 62 -----A-----E-VSI-----N 67
Db 420 SQNRGAUGVAFYHQLORVFTVQNVKVSVDINGLNIQEVLVNSVETPENLAVDWVN 479
QY 68 -----E-----DG---TL-----72
Db 480 KIVLVETKVNRIKIDWNLDSYRVTLITENLCHPRGIAVDPTVGYLFFSDWESLSGEPKLE 539
QY 73 -F-DG-R-----P-----IE-----S 80
Db 540 RAFMDGSNRKDLVKTLGNPAGVTLDMISKRVYVWDSRFDYIETVYDGIQKRTVHGG 599
QY 81 L-----SL-I-D---AVM-----88
Db 600 LIHPFGVSLFEGQVFTDTKMAVLKANKFTETNPQVYQASLRPYGVTVYHSLRQPYA 659
QY 89 -P-D-----V-V-----Q-T-R---Q-----Q-AY 99
Db 660 TNPCKDNGGCEQCVLSHRTDNDGLGFRCKCTFGFOLDTDERHACIAVQNLFISSQVAI 719
QY 100 R-----D-K-----L-AQ-----Q-----Q-A 108
Db 720 RGIPFTLSTQEDVMVPSVGNPSFFVGIDFDAQDSTIFFSDMSKHMIFKQIDGTGREILA 779
QY 109 A-----A-----A-----A-----A-- 112
Db 780 ANRVENVESLAFDWISKVLTWDSHYKSI SVMRLADKTRRTVVOYLNPNRSVVHPFAGY 839
QY 113 -----A-----A-----AA-----S-- 119
Db 840 LFPTDMFRPAKIMRAWSDGSHLLPVNTILGWPNGLAIDWAASRLYWDVAFDKIEHSTF 899
QY 120 -----Q-----120
Db 900 DGLDRRLGHIEQWTHPFGLAI FGEHLFFTDRWLGAIRVRKADGGEMTVIRSGIAYILH 959
QY 121 -----Q-GS-A-----KNG-----EN-----TA-N-----G-----133
Db 960 LKSYDNIQTSNACNPTHPNGDCHSFPPVFNQVRCPCPYGMRLASNLHLCBGDPTN 1019
QY 134 E---E-----N-----GA---H-----T-----140
Db 1020 EPTEQCGLSFPCKNGRCVPNYLDCGVDDCHNSDEQLCGTLNNTCSSAFTCGHGEC 1079
QY 141 I-A-----N-----H-----T-DME-----V-D---GD---154
Db 1080 IPAHWRCDKNDKCDVGSDEHNCPTHPASCLDTQYTCNDHQICISKNWVCDTNDNCGDSD 1139
QY 155 -----VEI---P---PN-----K-A-----162
Db 1140 EKNCNSTETQPSQFNCNHRICIDLSPVCDGDKDVGDSDEVCVNLNCTASQPKCASGDK 1199
QY 163 -----V-----VL-----R-G-H-----E-----169
Db 1200 CIGVTRCDGVFDCSDNSDEAGCPTRPFGMCHSDEFCQEDGICIPNFWECDGHPDCLYG 1259
QY 170 S-E-----V-----F-----IC---AW-----N-----P-----179
Db 1260 SDEHNACVPKTPSSYFHCNCGNCIHRALWICDRDNDGCDMSDEKDCPTQFPFCPSMOWQC 1319

QY 180 -----V-----SD-----182
Db 1320 LGHNICVNLVDCGIFDCPNGTDESPLCNGNSCSDFNCGCTHECVQEPFGAKCLPLGF 1379
QY 183 LLA-----S-----GS-----G-----189
Db 1380 LLANDSKTCEDIDECIDILGSCSQHCYNMRGSRFCSDTYMLBESDGRCTCKVTASESILL 1439
QY 190 -----DS-T-----A-----RI-----W---N---197
Db 1440 VASQNKIADSVTSQVHNIYSLVNGSYIVAVDFDSISGRIFWSDATQGTWSAFONGTD 1499
QY 198 -----LSE-----N---S---T---S---GS-----T-Q-LVL 211
Db 1500 RRVVFDSSILTTETIAIDWGRNLYWTDYALETIEVSKIDGSHRTVLISKNLINPRGLAL 1559
QY 212 --R-----H-----213
Db 1560 DPMNEHLLFWSDWGHHPRIERASMDGSMRTVIQDKIFWPCGLTIDYPNRLLYFMDSYL 1619
QY 214 -----C-----IR-----E-----GG-----219
Db 1620 DYMDFCDYNGHHRQVIASDLIIIRHPYALTLFEDSVYWTDRTRVRMRANKMHGQNSV 1679
QY 220 -----Q-----220
Db 1680 MYNIQWPLGIVAVHPKQPNVNPFCASRCSHLCILSSQGPFPYSCVPSGWSLSPDLLN 1739
QY 221 --D-----V-----P---SN-K---DV-----228
Db 1740 CLRDDQPFLLITVRQHIIFGISLNPVKSDAMVPIAGIQLDGLVDFEDDABQIYVWENPG 1799
QY 229 -----T-----S---LDW-----N---S-E---G-----T 238
Db 1800 EHRVKTGNTRTVFASISMGVSMNLALDWSIRNLSTYNTPTQSIETVLTHGDIRYKT 1859
QY 239 LLA---T---G-----S-----Y-----DG-----F-----248
Db 1860 LIANDGTALGVGPPIGITVDPPARKLYWSDQDTSVPKAKIASANMDGTSVKTLFTGNLE 1919
QY 249 -----A-----R-I-----W-----T---253
Db 1920 HLECVTLDEEOKLYWAVTGRGVIERGNDGTDMLVHQLSHPWGIAVHDSFLYTYDEQ 1979
QY 254 --K-D---G-N---L---ASTL-G-Q---H-----K-----266
Db 1980 YEIVERVKATGANKIVLRDNPVNLRLQVYTHRRNAESSNGCSNNMNACQICLPVPGG 2039
QY 267 -----G-----P-----I---PALK-----273
Db 2040 LFSACATGKLPNDRNSCPYNSFIVMSLSAIRGFSLELSDHSETWVPVAGQGRNALH 2099
QY 274 -----W-----N---K-----K-----GN- 279
Db 2100 VDVVSSGFIWCDPSSSVASDNAIRRIKPDGSSLANIVTHGIGENGVRGVIADWVAGNL 2159
QY 280 -----L-S-----A---G---VD-----F-----I-----281
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVDMPRHIVVDPKNRYLFWADYQORPKIERSF 2219
QY 282 -----L-S-----A---G---VD-----287
Db 2220 LDCNTRTVLSEGIVTTPRGLAVDRSDGYVYVWDDSLDIATIRINGENSEVIRYSGRYPT 2279
QY 288 -----K-----T---T-I---I-W-----D-----294
Db 2280 PYGITVPENSIWDRNLKKI FOASKEPENTEPPTVIRDNINWLRDVTTFDKQVQPSPA 2339
QY 295 -----A---HT-----G-----E-----A-----300
Db 2340 EVNNNPCLNNGGCHLCFALPGLHTPKDCAPGTLQSDKGNCAISTENFLIPALSNSLR 2399

301 QY -----K-----Q-----Q-F-----P-305
2400 Db SLHLDPENHSPFQTINVERTVMSLDYSDVSDRIYFTQNLASGVGQISVATLSSGHTPT 2459
306 QY -----F----- 306
2460 Db VIASGIGTAGIAFWITRIIYSDYLNQMINSGAEDSNRTVIARVPRAIPLVDPQCG 2519
307 QY -----H-----S--AP--ALD----- 313
2520 Db YLYWADWTHAKIERATLGNFRVPIVNSLVMPGSLTLDYEDLLYVWDASLQRIERST 2579
314 QY -----VD-----W-----Q-----S-----N----- 319
2580 Db LTGVREVIYNAVAHAFGLTYGYIYWTDLTYQRIYRANKYDGSQGIAMTTLNLSQPRG 2639
320 QY -----NT-----F----- 322
2640 Db INTVWKQKQCNPCQFNGCGSHICAPGPGAECQCPHEGNWYANNRKHCIVDNGER 2699
323 QY -----AS-----CS-----TD-M-----C-IH-----V----- 333
2700 Db CGASSFTCSNGRCISEEWMKCDNDNDCGSDMESVCAHTCSTPTAFTCANGRCVQYSYR 2759
334 QY -----G-----QD-----R-P--I----- 342
2760 Db CDYNDGCGSDGAGCLFRONATTEPMCMNRRCIPREFICNGVDNCHDNTSDCKNCPD 2819
343 QY -----K----- 346
2820 Db RTCQSGYTKCHNSNICIPRVVLCDGNDGNDSDENFTYCTTHTCSSEFQACASGRCIPO 2879
347 QY -----GH-----T----- 349
2880 Db HWYCDQETCFDASDEPASCBSHRTCLADEPKDGGRCIPSEWICDGDNDGMSDEDK 2939
350 QY -----N-----E-----VNA-----IK--W----- 357
2940 Db RHQONQNCSDSEFLCVNDRPDRBCIPQSVWCDGVDCDGDYDENQNCRTTCSENEFT 2999
358 QY ----- 357
3000 Db CGYGLCIPKIFCRDHNDCGYSDERGCLYQTCQOQFTCQNGRCISKTEFVCDENDCDG 3059
358 QY -----D-----PT-----G-----N-----L----- 363
3060 Db GSELMHLCHTPEPTCPHFKDNGRCIEMMKLCNHLDCDLDNSDEKGCINECHDPSI 3119
364 QY -----LAS--CS-----D-D--M----- 371
3120 Db SGCDHNCITLTSFYCSCRFGYKMSDKRTCVDIDECTEMPFVCSQKCNVIGSYICKCA 3179
372 QY -----T-----L----- 374
3180 Db PGLYREFDGTQRQNSNIEPIYFNRYLNLTDIGFYSLILEGLDNVVALDFDVEK 3239
375 QY -----I----- 377
3240 Db RLYWIDTQRQVIERMFLNKTNETIINHLPAESLAVDWVSRKLYWLDARLDGLFVSDDL 3299
378 QY -----M-KQ--D--N--C-----VHDLQO-----H-----NK-392
3300 Db NGHRRMLAQHCVDANTTFCDNPRGLAH--POGYLYWADWGHRAVIGRVGMDGTNKS 3357
393 QY -----E-----I-YT-----IK----- 398
3358 Db VIISTKLEWPNGITIDVTLNLLYWADAHGLGYIEYSDLEHGHRTVYDGPALPHPAITIFE 3417
399 QY -----W-----S-----P-T--G--403
3418 Db DTIYWDNTNTRTBKNGKDGNSRQTLVNTTHRPFDIHVYHPYQPIVSNPCGTFNNGCS 3477
404 QY -----PG-----T-----NN----- 408

3478 Db HLCLKPGKGFTCECPDDFRTLQLSGSTYCMPCMSSTOFLCANNEKCIPIWMKCDGQKD 3537
409 QY -----P-----NA--N-----LML-----AS 417
3538 Db CSDGSELALCPQRCFLQOFCSDNGNCTSPQTLNAHNCPPDGSDEDL-LCENHHDS 3596
418 QY -----A-----S-----F-----DS-----T-----VR-----L--W--DV 429
3597 Db NEWQCANKECIESWCQDTFNDCEDNSDEDSHCASRTCPQCFRANGRCIPOAMKCDV 3656
430 QY -----DRG-----I--CI--H-----T-L----- 438
3657 Db DNDGDSDEPIEECMSSAHLCDNFTEFSCKTNYRCIPKWAVGCVDDCRDSDGECCE 3716
439 QY -----T-----K--H-----Q-----E--P-----V----- 445
3717 Db RTCHPVGDPRCKNHHCIPLRWQCDGQNDGNDSCAPRECTESEFRVCNOQCIPSRW 3776
446 QY ----- 450
3777 Db ICDHYNDGNSDERDCENMRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836
451 QY -----SPDGRYL-AS-----GS-----F----- 462
3837 Db PDGAYCATMECKNHVCIPPYWKCDGDDCGSDGDEELHLCLDVPCNSPNRFRCDNNR 3895
463 QY -----D-----K-----CV-H--I----- 468
3896 Db CIYSHEVNGVDDCGDGTDETEHCKRTPKPCTEYKCGNGHCIPHONVDDADDGCD 3955
469 QY -----W-----N-----T-----Q--T-----G-----A-----L-V----- 477
3956 Db WSELGCKNGKERTCAENICEQNCITQNEGFTICSTAGFTNVFDRTSCLDINECEQFG 4015
478 QY -----H-----SY-----R----- 481
4016 Db TCPQHCRNTKGSYECVACDGTSMSTRPGKRCAAEGSPILLPLPDNVIRKYNLSERFS 4075
482 QY -----G-----T-----G-----G-----I--FE-----V-C--490
4076 Db EYLQDEYIQAVDYDWDPKDIGLSVVYTVRGESRFGAIKRAYIINFESGRNLLVQEV 4135
491 QY -----W-----N-----AA----- 494
4136 Db LKLYVMQPDGLAVDWGRHYIWSDVKNKRIEVALKLDGRVKNLSTDLDPQAAIAVNP 4195
495 QY -----K--V--498
4196 Db LGLMFWTDWKEPKIESAWMNGEDRNILVFDLWPTGLSIDYLNNDRIYWSDFKEDVIE 4255
499 QY -----G-----A-SA--S-----D-----G----- 505
4256 Db TIKYDGTDRRVIKAEAMNPYSLDIFEDQLYWISKEKEVWKNQKFGGKKEKTLVNPWL 4315
506 QY -----SV--C--V--L-----D----- 511
4316 Db TVRIFHQLRYNKSVPNLKQICSHLCLLRPGGYSCACPGSSFIESTTECDAAIELPI 4375
512 QY -----L-----R-----K 514
4376 Db NLPPCKRMHGGNCYFDETDLPK 4398

RESULT 6

US-08-652-877-88
; Sequence 88, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Raak, Lars
; APPLICANT: Crumley, Gregg R.

```

Query Match      70.1%; Score 2566.6; DB 3; Length 4655;
Best Local Similarity 9.8%; Pred. No. 5.1e-32;
Matches 433; Conservative 57; Mismatches 19; Indels 3894; Gaps 381;

Qy      1 M-----S-----I-----S-SDE- 7
      | |
Db      1 MDRGPAAVACTLLALVACLAPASGQECDSAHFRCGSGHCIPADWRCDGTKCSDDADEI 60
      | |
Qy      8 ---VN---F-----L---V-----12
      | |
Db      61 GCAVVTQQGQYFKQSEGCIPSSWVCDQDQDCDGDSDERQDCSQSTCSSHQTCSNGQC 120
      | |
Qy      13 ---YR-----Y---L-----Q-----E--S 19
      ||
Db      121 IPSEYRCDHVRDCPDGADENDCQYPTCEQLTCDNGACVNTSQKCDWKVDCRDSSEINCT 180
      | |
Qy      20 ---G-GS-----H-----S-A-F-T---F--G-I-----30
      || |
Db      181 EICLHNEFSCNGGECIPRAYVCDHDNDQDGSDEHACNYPCTGGYQFTCPSGRCIYQNWV 240
      | |
Qy      31 -----K-----S---H-----IS-----Q--SNI 39
      | |
Db      241 CDGEDCKNDGDEGCSGPHDVHKCSPREWSCPESGRCISYKVCDCGILDCPREDEN- 299
      | |
Qy      40 N---G-----AL-----VP-----PAAL I---S-----I-I---53
      | |
Db      300 NTSGTGYCSMTLCSALNCOYCHETPYGGACFCPPGVIINHDSRTCEVFDDCOIGWICD 359

```

QY 190 ---DS-T---A---RI---W---N--- 197
Db 1440 VASONKIIADSVTSQVHNIYSLVENGSIYVAVDFDSISGRIFWSDATQGTWGAFOGTD 1499
QY 198 ---LSE---N---S---T---S---GS---T---Q---LVL 211
Db 1500 RRVVDFSSIIITETIAIDWGRNLYWTDYALEITEVSKIDGSHRTVLISKNLNPRLAL 1559
QY 212 ---R---H--- 213
Db 1560 DPMNEHLLFWSDGHHPRIERASMDGSMRTIVQDKIFWPCGLTIDPYNRLLYFMDSYL 1619
QY 214 ---C---IR---E--- 219
Db 1620 DYMDFCYNGHRRROVITASLLIRHPYALTFLFDSVYVTRATRRVRANKWEGNQSVV 1679
QY 220 ---Q--- 220
Db 1680 MYNIOWPLGIVAVHPKQPNVPCAFSRCSHLCLLSQGPHFYSCVPCSGWSLSPLLLN 1739
QY 221 ---D---V---P---SN---K---DV--- 228
Db 1740 CLRDDOPFLITVRQHIIFGISLNPVKNSDAMVPIAGIONGLDVEFDDAQIYIYVENPG 1799
QY 229 ---T---S---LDM---N---S---E---G---T 238
Db 1800 EIHRTVDGNRTVFAISVMGFSMNLALOWISRLNSTNPTQTSIEVLTGLHGDIRYKT 1859
QY 239 LLA---T---G---S---Y---DG---F--- 248
Db 1860 LIANDGTALGVGPIGITVDPARGKLYWSDQDTSVGPAKIASANMDGTSVKTLFTGNLE 1919
QY 249 ---A---R---I---W---T--- 253
Db 1920 HLECVTLDIEQKLYMAVTCRGVIERGNVDGDMILVHOLSHPWGIAVHDSFLYTTDEQ 1979
QY 254 ---K---D---G---N---L---ASTL---Q---O---H---K--- 266
Db 1980 YEIVERTKATGANKIVLRDNVPLRLQVYHRRNAEASSNGCSNNWACQQICLPVPGG 2039
QY 267 ---G---P---I---FALK--- 273
Db 2040 LFSCACATGKFLNPDNRCSPYNSFIVVSMLSAIRGFSLESDHSETMVPVAGQGRNALH 2099
QY 274 ---W---N---K---K---GN--- 279
Db 2100 VDVVSSGFTYWCDFSSSVASDNAIRRIKPDGSSLMNIVTHGIGENGVRGIAVDWVAGNL 2159
QY 280 ---I---F--- 281
Db 2160 YFTNAFVSELTIEVLRINTYRRVLLKVTVDMPRHIVVDPNRYLFWADYQRPKIERSF 2219
QY 282 ---L---S---A---G---VD--- 287
Db 2220 LDCTNRTVLVSEGITVPRGLAVDRSDGVYVWDDSLDIARIRINGENSEVIRYGRYPT 2279
QY 288 ---K---T---T---I---I---W---D--- 294
Db 2280 PYGITVFENSIWVDRNLKXIFQASKEPENTPEPTVIRDNINLWLDVTIFDKQVQPRSPA 2339
QY 295 ---A---HT---G---E---A--- 300
Db 2340 EVNNPCLNNGGSHLCFALPGLHTPKCDCAFGLTQSDGKNCAISTENFLIFALNSLR 2399
QY 301 ---K---Q---Q---F---P--- 305
Db 2400 SLHLDPNHSPFPQTIWVTRVMSLDYSDVSDRIYFTQNLASGVQIYATLSSGIHTPT 2459
QY 306 ---F--- 306
Db 2460 VIASGIGTAGIADFTWTRRIYYSYDLNQMINSMAEDGSNRTVIARVPKPRAIVLDPCCQ 2519
QY 307 ---H---S---AP---ALD--- 313

Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSSLVMPSSGLTLDYEBEDLLYWDVDSLQRIERST 2579
QY 314 ---VD---W---Q---S---N--- 319
Db 2580 LTGVDRREVIVNAAHAFGLTLYGQYIYWTDLTYQRIYRANKYDGSQIAMIATNLLSQPRG 2639
QY 320 ---NT---F--- 322
Db 2640 INTVKNQKQCNPCPEQNGGSHICAPGNGABEQCEHGWYLANNRKHCIUVNGER 2699
QY 323 ---AS---CS---TD---M---C---IH---V--- 333
Db 2700 CGASSFTCSNGRCISBEWKCDNDNDGDSDEMSVCALHTCSPTAFTCANGRCVQYSYR 2759
QY 334 CKL---G---QD---R---P---I--- 342
Db 2760 CDYINDGDSDEAGCLFRDCNATTFEFCNNRRRCIPREFICNGVDNCHDNNTSDEKNCPD 2819
QY 343 ---K---T---FO--- 346
Db 2820 RITCSGYTTHCHNSNICIPRYLDCGNDNDGNSDENPTYCTHTCSSEFQASGRICIPQ 2879
QY 347 ---GH---T--- 349
Db 2880 HWYCDQETDCFDASDEPASCGHSERTCLADEPKDCGRCIPSEWICDGDNDGDMDEDK 2939
QY 350 ---N---E---VNA---IK---W--- 357
Db 2940 RHOCQONQSDSEFLCVNDRPPDRRCIPOSWVCGDGDVDCDGDYDENONCTRRCTSENEFT 2999
QY 358 --- 357
Db 3000 CGYGLCIPKIFRCDRINDCGYSDERGCLYQTCQOQFTCQNGRCISKTFVCDENDCGD 3059
QY 358 ---D---PT---G---N---L--- 363
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLDNDSEKGGGINECHDPSI 3119
QY 364 ---LAS---CS---D---D---M--- 371
Db 3120 SGCDHNCTDLTSFYCSRPGYKLMDSKRTCDVIDECTEMPFVCSOKCENIVGSIYCKCA 3179
QY 372 ---T---L--- 374
Db 3180 PGYLREPDGKTCEQNSNIEPYLIFSNRYLRLNLTIDGYFYSLLILEGLDNVVALDFRVEK 3239
QY 375 ---I---W---S--- 377
Db 3240 RLYWIDTQOVIERMFLNKTNETIINHRLPAEASLAVDWVSRKLYWLDARLDGLFVSDL 3299
QY 378 ---M---KQ---D---N---C---VHDLOO---H---NK--- 392
Db 3300 NGGHRMLAQHCVDANNTFCFDPNPRGLALH---PQYGLYWADWGHRAIYGRVGMGTNKS 3357
QY 393 ---E---I---YT---IK--- 398
Db 3358 VIISTKLEWPNGTIDYTNLLYWAHAHLGYIEYSDLEGGHHRHTVVDGALPHFAITIFE 3417
QY 399 ---W---S---P---T---G--- 403
Db 3418 DTIYTDWNRTRTVEKNGKYDGSNRQTLVNTTHRPFDIHVVHYRQPIVSNPCGTNNGCS 3477
QY 404 ---PG---T---NN--- 408
Db 3478 HLCLIXPGGKFTCECPDDPRTLQLSGSTVCMPCWCSSTQFLCANNEKCIPIWKCQDQKD 3537
QY 409 ---P---NA---N---LML---AS 417
Db 3538 CSDGSDALALCPORFCLGQFCQDGNCTSPQTLCAHQNCPDGSDDEDL-L-CENHHCHDS 3596
QY 418 ---A---S---F---DS---T---VR---L---W---DV 429

Db 3597 NEWQANKRCIPESWQCDTTFNDCEDNSDESSHCASRTCRPGQFRANGRCRIPQAWKCDV 3656
QY 430 --DRG-----I--CI--H-----T--L----- 438
Db 3657 DNDGDSDEPIECSSAHLCDNFTEFSKTYRCIPKAWCVGDDCDNDSDGQCEE 3716
QY 439 -T-----K--H-----Q-----E--P-----V----- 445
Db 3717 RTCHPVGDFRCKNHCIPLRWQDQNDGNDSENCAPRECTESEFRVCVQOCIPSRW 3776
QY 446 -----Y-----S--V-----A-----F 450
Db 3777 ICDHYNDGNSDRDCMRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836
QY 451 SPGRYL-AS-----GS-----F----- 462
Db 3837 -PDGATQATMFCKKHVCIPPYWKDGDGDDGDSDEELHCLDVPKNSPFRCDNR 3895
QY 463 -----D-----K-----CV-H-I----- 468
Db 3896 CIYSHEVCNGVDDGDTDEBHCRTKPKCTEYKCGNGHCIPHDNVCDADDDCGD 3955
QY 469 W-----N-----Q--T-----G-----A-----L-V----- 477
Db 3956 WSDLGCKGKERTCAENICEONCTQNEGFFCSCCTAGFTNVFDRTSCLDINECEQFG 4015
QY 478 -----H-----SV-----R----- 481
Db 4016 TCPQHCRNTKGSYECVACDGTSMDSRDPGRKCAABGSSPLLLLPDNRVIRKYNLSERFS 4075
QY 482 -----G-----T-----G-----G-----I--FE-----V-C-- 490
Db 4076 EYLQDEEYIOAVDYDMDPDIGLSVYVTVRGESRFGALKRAVINFESGRNLLVQEV 4135
QY 491 -----W-----N-----AA----- 494
Db 4136 LKLYVMQPDGIAVWVRHRYNSDVKNKRIEVAKLDRYRKWLSTDLDPAAIAVNP 4195
QY 495 -----G-----G-----D-K-V-- 498
Db 4196 LGLMFTWDMGKPKXSAMNGSDRNLVLPEDLGMPTGLSIDYLNNDRIYWSDFKSDVIE 4255
QY 499 -----G-----A-SA-----S-----D-----G----- 505
Db 4256 TIKYDGTDRVIAKEANPYSLDIFEDQLYWISKEGEVWKQNGKKEKTLVNPWL 4315
QY 506 -----SV-----C-V-----L-----D----- 511
Db 4316 TVRIFHQLRYNKSVPNLCKQICSHLCLLRPGGYSCACPOGSSFIEGSTTECDAAIELPI 4375
QY 512 -L-----R-----K 514
Db 4376 NLPPPCRMHGGNCYFDETDLPK 4398

RESULT 7

US-08-652-877-90

; Sequence 90, Application US/08652877

; Patent No. 6187548

; GENERAL INFORMATION:

; APPLICANT: Akerstrom, Goran

; APPLICANT: Juhlin, Claes

; APPLICANT: Rask, Lars

; APPLICANT: Crumley, Gregg R.

; APPLICANT: Morse, Clarence C.

; APPLICANT: Murray, Edward M.

; APPLICANT: Hjalms, Goran

; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments

; TITLE OF INVENTION: Thereof and DNA Encoding Same

; NUMBER OF SEQUENCES: 106

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Rhone-Poulenc Rorer Inc.

; STREET: 500 Arcola Rd., 3C43

; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426-0107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.5.1
; SOFTWARE: Word 6.0 (Patentin)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,877
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/15203
; FILING DATE: 22-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,836
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,314
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitzky, Martin
; REGISTRATION NUMBER: 29,699
; REFERENCE/DOCKET NUMBER: A1355E-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-454-3816
; TELEFAX: 610-454-3808
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4655 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-652-877-90

Query Match 70.1%; Score 2566.6; DB 3; Length 4655;

Best Local Similarity 9.8%; Pred. No. 5.1e-32;

Matches 433; Conservative 57; Mismatches 19; Indels 3894; Gaps 381;

QY 1 M-----S-----I-----S--SDE- 7
Db 1 MDRGPAVACTLLALVALCLAPASGQECDSAHFRGSGHCIPADWRCDGTGKDCSDDADEI 60
QY 8 ---VN---P---L---V----- 12
Db 61 GCAVVTCCQGYFKQSEGCIPSSWVCDQDQDCDGSDEQDCSQSTCSSHQITCSNGQC 120
QY 13 ---YR---Y---L---Q-----Q-----E--S 19
Db 121 IPSEYRCDHVRDCPDGADENDCOYPTCEQLTCDNGACYNTSOKCDWKVDCRSDSDEINCT 180
QY 20 ---G-FS-----H-----S--A--F--T-----F--G--I----- 30
Db 181 EICLHNEFSGNGECIPRAVVCDDHNDCCDGSDEHACNYPTCGGYQFTCPSGRCIYQNWV 240
QY 31 ---K---S-----H-----IS-----Q--SNI 39
Db 241 CDGEDCKDNGDBDGCESGPHDVHKCSPREWSCPSGRCISIIYKVCDDGILDGCPGREDEN- 299
QY 40 N--G-----AL-----VP-----PAALI-----S-----I--I-- 53
Db 300 NTSTGKCSMTLCSALNCYQCHETPYGGACFCFPGYIINHNSRVCVEFDQCIQWICD 359
QY 54 QK-----G--L--QY-----VE 61
Db 360 QKCESRGRHLCHCEBGYILERQYCKANDSPGEASIIFNSGRDLLIGDIHGRSFRILVE 419
QY 62 -----A-----E--VSI-----N 67
Db 420 SQNRGVAVGVAFHLYLQRVFWTDTQNVKVFSDINGLNIOEVLNVSFETPENLAVDWNV 479

Qy 68 ---E---DG---TL-----72
Db 480 KIYLVETKVRIDMWNLDGSRVTLITENLGHPRGIAVPTVGYLFFSDWESLSGEPKLE 539
Qy 73 -P-DG-R-----P-----IE-----S 80
Db 540 RAFMDGSRNRKDLVKTKLGPAGVTLDMISKRVYVWDSRFDYIETVTVYDGIQRKTVVHGS 599
Qy 81 L-----SL-----I---D---AVM-----88
Db 600 LIPHPGVSLEFGQVFTDTWKAVLKANKFETNPQVYQASLRPYGVTVYHSLROPYA 659
Qy 89 -P-D-----V-V-----Q-T-R-----Q-AY 99
Db 660 TNPCKDNNGGCEQCVLSHRTDNDGLGRCKTFGFOLDTDERHCIAVQNFIFSSQVAI 719
Qy 100 R-----D-K-----L-AQ-----Q-----A 108
Db 720 RGIPFTLSTQEDVMVPVSGNPSFVFGIDFDAQDSTIFFSDMSKHMIFKQKIDGTGREILA 779
Qy 109 A-----A-----A-----A-----A--112
Db 780 ANRVENVESLAFDWISKNLWYTDSHYKSI SVMRLADKTRTVVQYLNPNRSVVVHPAGY 839
Qy 113 -----A-----A-----AA-----S--119
Db 840 LFFTDWFRPAKIMPAWSGSHLLPVNTILGWPNGLAIDWAASRLYWVDAYFDKIEHSTF 899
Qy 120 -----Q-----Q-----Q-----Q-----120
Db 900 DGLDRRLGHIEQTHPFLAIFGEHLFFTDWRLGAIIRVKADGGEMTVIRSGIAVILH 959
Qy 121 -----Q-GS-A-----KNG-----EN-----TA-N-----G-----133
Db 960 LKSYDNIQTGSNACNQPHTPGDCSHFCFPVFNQVCGPYGMRLASNHLTCBGDPTN 1019
Qy 134 E-E-----N-----GA-H-----T-----140
Db 1020 EPPTQCGLFSPCKNGRCVNYLDCGVDDCHDNDSEQLCGTLNNTCSSAFTCGHGEC 1079
Qy 141 I-A-----N-----H-----T-DMME-----V-D-----GD-----154
Db 1080 IPAHRWCDCRNDKVDGSDHNCPTHAPASCLDTQYTCDNHCQISKWVCDTNDGCGSD 1139
Qy 155 -----VEI-P-----PN-----K-A-----162
Db 1140 EKNCNSTETCQPSQFNCNHRCDLSFVCDGDKDCVDGSDVGVCLNCTASQPKASGDK 1199
Qy 163 ---V-----VL-----R-G-H-----E-----169
Db 1200 CIGVTRCDGVDCSDNSDEAGCPTRPFGMCHSDFEQCEGICIPNFWCEDGHPDCLYG 1259
Qy 170 S-E-----V-----F-----IC-AW-----N-----P-----179
Db 1260 SDEHNACVPKTCPSYFHCNCGNCIHRALWLCDRNDGDMSEDKDCPTQFRCPSQWQC 1319
Qy 180 -----V-----SD-----182
Db 1320 LGHNICVNLVSVCDGIFDCPNGTDESPLCNGNSCDFNGGCTHECVQEPFGAKCLCPLGF 1379
Qy 183 LLA-----S-----GS-----G-----189
Db 1380 LLANDSKTEBIDBCDILGSCSQHCYNMRGSRFCSDTGYMLBDDGRTCKVTASESILLL 1439
Qy 190 -----DS-T-----A-----RI-----W-----N-----197
Db 1440 VASQNKIADSVTSQVHNIYSLVNGSYIIVAVDFDSISGRIFWSDATQGTWSAFQNGTD 1499
Qy 198 -----LSE-----N-S-----T-S-----GS-----T-Q-LVL 211
Db 1500 RRVVFDSSIIILTETIADWGRNDLYWTDYALETIEVSKIDGSHRTVLISKNLNPRGLAL 1559
Qy 212 --R-----H-----213

Db 1560 DPRMNEHLLFWSDGWHHPRIERASMDGSMRTVIVQDKIWFPCGLTIDYPNRLLYFMDSYL 1619
Qy 214 ---C-----IR-----E-----GG-----219
Db 1620 DYMDFCDYNGHRRORVIASDLIIRHPYALTLPEDSVYWTDRATRRVVRANKMWHGQSVV 1679
Qy 220 -----O-----220
Db 1680 MYNIQWPLGIVAVHPKQPNVNCAPFRCRSHLCLSSQGPHPFYCVPCSGWLSLSPDLLN 1739
Qy 221 ---D-----V-----P-SN-K-----DV-----228
Db 1740 CLRDDQFLITVRQHIIIFGILNPEVKSDAMVPIAGIQLGLDVEFPDABQIYVWENPG 1799
Qy 229 ---T-----S-LDW-----N-S-E-G-----T 238
Db 1800 EIHVRKTDGTRNRTVPASISMVGFSMNLALDWISRLNLYSTNPTQSIETLHGDIRYRK 1859
Qy 239 LLA---T-G-----S-Y-----DG-----F-----248
Db 1860 LIANDGTALGVPPIGITVDPARGKLYWSDQGTDSVPKIASANMDGTSVKTLFTGNLE 1919
Qy 249 -----A-----R-I-----W-----T-----253
Db 1920 HLECVTLDIEEQKLYWAVTGRGVIERGNVDGTRMILVHQLSHPMGIAVHDSFLYVDEQ 1979
Qy 254 -K-D-G-N-L-ASTL-G-Q-H-----K-----266
Db 1980 YEVIERVKATGANKIVLDRDNVNLRLGLQVYHRRNAESNGCSNNMACQICLPVPGG 2039
Qy 267 ---G-----P-----I-FALK-----273
Db 2040 LFSCACATGKLPNDRSCSPNSFIVVMSLGAIRGFSLESDHSETMPVACQGRNALH 2099
Qy 274 ---W-----N-----K-----K-----GN-279
Db 2100 VDVVSSGFIYMCDFSSVASDIAIRIKPDGSSLMNIVTHGIGENGVRGIAVDWAGNL 2159
Qy 280 -----F-----I-----281
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVMRPHRVDPKRYLFWADYQORPKIERSF 2219
Qy 282 ---L-S-----A-G---VD-----287
Db 2220 LDCNTRVLVSEGIPTPRGLAVDRSDGYVWDDSLDIATIRINGENSEVIRYGRSPT 2279
Qy 288 ---K---T---T-I---I-W-----D-----294
Db 2280 PYGITVFENSIWVDRNLKKIIOASKEPEPTEPTVIRDNINWLDRDVTIFDKQVQPRSPA 2339
Qy 295 ---A---HT-----G-----E-----A-----300
Db 2340 EVNNPCLENNCGCSHLCPALPGLHTPKDCAFGLQSDGKNCAISTENFLIPALSNSLR 2399
Qy 301 ---K---Q-----Q-F-----P-305
Db 2400 SLHLDPENHSPPQTINVERTVMSLDYSDSRIYFTQNLASGVQIATLSSGHTPT 2459
Qy 306 -----F-----306
Db 2460 VIASGIGTAGIAFDWITRRIYVDYNQWINSMAEDSGSNRTVIARVPKPRAIVLDPQCG 2519
Qy 307 ---H-----S-AP---ALD-----313
Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSLWPSGLTLDYEEDLLYVWDASLQRIERST 2579
Qy 314 ---VD-----W---Q-----S---N-----319
Db 2580 LIGVDREVIVNAVHAFGLTYLQYIYWTDLTYQRIYRANKYDGSQCIAMTNLLSQPRG 2639
Qy 320 -NT-----F-----322

Db 2640 INTVKNQKQCCNPNCEQFNGGCSHCAPGNGABEQCPHEGNYLANNRKHICVNDGER 2699
Qy 323 --AS--CS-----TD-M--C-IH-----V----- 333
Db 2700 CGASSFTCSNGRCISEBWKCDNDNDGDSDEMSVCAHTCSPTAFTCANGRCVQYSYR 2759
Qy 334 CKL-----G-QD-----R-P-I----- 342
Db 2760 CDYVDCGDSDEAGLCFRDCNATTEPMCNRRCPREFICNGVDNCHDNTSDEKNCPD 2819
Qy 343 -----K-----T-----PQ----- 346
Db 2820 RTCOSYTKCHNSNICIPRYLDCGNDGNSDENPTYCTTHTCSSEFQCSAGRCIPQ 2879
Qy 347 -----GH--T----- 349
Db 2880 HWYCDQETCFDASDEPASGHSERTCLADEFKDGGRCIPSEWICGDNDCGMSDEDK 2939
Qy 350 -----N--E--VNA-----IK--W----- 357
Db 2940 RHQONQNCSDSEFLCVNDRPPDRRCIPQSWVCDGDVDCDGYDENQNCRTTRTCSNEFT 2999
Qy 358 -----PT-----G-----N--L----- 357
Db 3000 CGYGLCIPKIFCRDRHNDGYSDBRGCLYQTCQOQFTCQNGRCISKTFVCDENDDCG 3059
Qy 358 --D--D-----PT-----G-----N--L----- 363
Db 3060 GSDLMHLCHTPTCPPEFKDNGRCIEMKLCNHLDCDLSDEKGGCINECHDPSI 3119
Qy 364 -----LAS--CS-----D-D--M----- 371
Db 3120 SGCDHNCTDLTSFYCSRGYKMSDKRTCVDIDECTEMPFVCSKCNVIGSYICKCA 3179
Qy 372 -----T-----L----- 374
Db 3180 PGYLRPDGKTCRQNSIEPIYFNRYLNLTDGYFYSYLEGLDNVVALDFRVEK 3239
Qy 375 -----I-----S-----W----- 377
Db 3240 RLYWIDTQVIERMFLNKTNETIINHRLPAESLAVDWWSRKLWLDARLDGLFVSDL 3299
Qy 378 -----M-KQ--D--N--C-----VHDLOQ-----H-----NK-- 392
Db 3300 NGHRRMLAQHCVDANNTFCFNDPRGLAH--PQYGLYWADGHRAYIGRVGMDGTNKS 3357
Qy 393 -----E-----I-YT-----IK----- 398
Db 3358 VIISTKLEWPNGITIDYTNLLYWAHAHLGYIEYSDLEGHHRHTVYDGLPHFPFAITFE 3417
Qy 399 -----W-----S-----P-T--G-- 403
Db 3418 DTIYTDWTRTVEKNGKYGSRQTLVNTTHRPDIHVHPYRQPIVSPCOTNNGCS 3477
Qy 404 -----PG-----T-----NN----- 408
Db 3478 HLCILPFGGKFTCECPDDPTLQSGSTYCMPCSSQFLCANNEKCIPIWKKCDGQXD 3537
Qy 409 -----P-----NA--N-----LML-----AS 417
Db 3538 CSDGDELALCPQRCFLGQFCSDGNCTSPOTLCNAHQCPDGSDEDL-LCENHCHDS 3596
Qy 418 -----A-----S-----P-----DS-----T-----VR-----L--N--DV 429
Db 3597 NEWQCANRCIPESWQCDTNDCEDESDSSHCASRTCPGQFRANGRCIPQAWKCDV 3656
Qy 430 --DRG-----I--CI--H-----T--L----- 438
Db 3657 DNDGDSHDSPIECMSAHLCDNFTFESCKTYRCIPKWAVCNVYDDCDNDSDEQCEE 3716
Qy 439 --T-----K--H-----Q-----E--P-----V----- 445
Db 3717 RTCHPVGDFRCNKHHCIPLRWQCDGQNDGNSDEENCAPECTSEFRVCVQCCIPSRW 3776

Qy 446 -----Y--S--V-----A-----F 450
Db 3777 ICDHYNDCGNSDERDCMRKTCHEPYFQCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836
Qy 451 SPDGRYL-AS-----GS-----F----- 462
Db 3837 -PDGAYCOATMECKKNHCIPPYWKCDGDDDCGDSDEELHLCIDVPCNSPNRFRCDNNR 3895
Qy 463 -----D-----K-----CV-H--I----- 468
Db 3896 CIYSHEVCNGVDDCGGTDTEHEHCKRTPKPCTEYKCGNGHCHPHDNDVCDADDCGD 3955
Qy 469 W-----N-----T-----Q--T-----G-----A-----L--V----- 477
Db 3956 WSDELCGNKGRKTCACENICEQNCQTLNCGGFCSCCTAGFETNVFDTSCLDINECEQFG 4015
Qy 478 -----H-----SY-----R----- 481
Db 4016 TCPQHCRNTKGSYECVACADGFTSMRPGKRCAAEGSSPLLLLPDNNVRIRKYNLSERFS 4075
Qy 482 -----G-----T-----G--G--G-----I--EE-----V--C-- 490
Db 4076 EYLQBEYIQAVDYDWDPEDIGLSVYVYVRGSGRFGAKRAYIPNFESGRNNLVQEV 4135
Qy 491 -----W-----N-----AA----- 494
Db 4136 LKLYVMQPDGIAVDWVGRHIYWSDVKNKRIEVAKLDGRYKRWLSTDLDOOPAAIAVNP 4195
Qy 495 -----G-----D-K--V-- 498
Db 4196 LGLMFTWKGPKLESAMNGEDRNILVFDLGMPTGLSIDYLNNDRIYWSDFKEDVIE 4255
Qy 499 -----G-----A-SA--S-----D-----G----- 505
Db 4256 TIKYDGTDRRVTAKEANNPYSLDIFEDQLYWISKEKEVWKNQKFGGKKEKTLVNPWL 4315
Qy 506 -----SV--C--V-----L-----D----- 511
Db 4316 TVVRIFHQLRYNKSVPNLCKQICSHLCILRPGYSCACPGQSSFIEGSTTECDAAIELPI 4375
Qy 512 -L-----R-----K 514
Db 4376 NLPPPCRCMHGNCYFDETDLPK 4398

RESULT 8

US-08-652-877-86
; Sequence 86, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Raak, Lars
; APPLICANT: Crumley, Gregg R.
; APPLICANT: Morse, Clarence C.
; APPLICANT: Murray, Edward M.
; APPLICANT: Hjalms, Goran
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
; TITLE OF INVENTION: Thereof and DNA Encoding Same
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd., 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426-0107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.5.1
; SOFTWARE: Word 6.0 (Patentin)

Db 1680 MYNIQWPLGIVAVHPKSPNSVNPFCARSRCSHLCLLSSQGHFHYSCVCPSPGWSLSPDLLN 1739
Qy 221 ---D---V---P---SN---K---DV--- 228
Db 1740 CLRDDQFLITVRQHIIIFIGISLPEVKSNDAMVPIAGIQLGDLVEFPDDASQIYIWWENPG 1799
Qy 229 ---T---S---LDW---N---S---E---G---T 238
Db 1800 EIHVRKTDGTRNVTFASISVMGSPMNLALDWISRLNLYSTNPTQSI EVLTLHGDIRVKT 1859
Qy 239 LLA---T---G---S---Y---DG---F--- 248
Db 1860 LIANDGALGVGPIGIVDPARKLYWSQDGSVPAKIASANMDGTSVKTLFTGNLE 1919
Qy 249 ---A---R---I---W---T--- 253
Db 1920 HLECVTLIDIEQKLYNAVTGRGVIERNVGDTRMILVHQLSHPWGLIAVHDSFLYTYDEQ 1979
Qy 254 ---K---D---G---N---L---ASTL---G---Q---H---K--- 266
Db 1980 YEVIERYDKATGANKIVLRDNVPLRGLOVYHRRNAEASSNGCSNNMNAQQICLPVPGG 2039
Qy 267 ---G---P---I---FALK--- 273
Db 2040 LFSCACATGFLAPDNRSRCSFYNSFIVVMSLSAIRGFSLSHDSETMVPVAGQGRNALH 2099
Qy 274 ---W---N---K---K--- 279
Db 2100 VDVVSSGFIYWCDFSSVASDNAIRIKPDGSSLMNIVTHGIGENGVRGIAVDWAGNL 2159
Qy 280 ---I---F---I--- 281
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVDMPRHIVVDPKRYLFWADYQRPKERSP 2219
Qy 282 ---L---S---A---G---VD--- 287
Db 2220 LDCTNRTVLVSEGIIVTPRGIAVDRSDGYVYVVDSDLDIARIINGENSEVIRYGRYPT 2279
Qy 288 ---K---T---T---I---I---W---D--- 294
Db 2280 PYGITVFENSIIWDRNLKIKIFOASKEPENTEPPTVIRDNINWLRDVTFDKQVQPRSPA 2339
Qy 295 ---A---HT---G---E---A--- 300
Db 2340 EVNNNPCLENGGCSHLCPALPGLHPKCDCAFGLTQSDGKNCAISTENFLIFALNSLR 2399
Qy 301 ---K---Q---Q---F---P--- 305
Db 2400 SLHLDPENHSPPQTINVERTVMSLDYSDRIYFTQNLASGVGQTSVATLSSGIHTPT 2459
Qy 306 ---F--- 306
Db 2460 VIASGIGTAGIADWITRIIYSDYLNQMINMAEDGNSRNTVIARVPKRAIVLDPCCQ 2519
Qy 307 ---H---S---AP---ALD--- 313
Db 2520 YLWADWDTHAKIERATLGNFRVPIVNSLVPSPGLTDYEDLLYWDASLQRIERST 2579
Qy 314 ---VD---W---Q---S---N--- 319
Db 2580 LTGVDRREVIVNAAHAFGLTYGYIYWTDLTYRIYRANKYDGSQGIAMTTLNLSQPRG 2639
Qy 320 ---NT---F--- 322
Db 2640 INTVVRKQKQCCNPNCEQFNGGCSHTCAPFNGAECQCPEHGNWYLANNRKHCIVDNGER 2699
Qy 323 ---AS---CS---TD---M---C---IH---V--- 333
Db 2700 CGASSFTCSNGRCISBEWKCDNDGCGSDMESVCAHTCSPTAFTCANGRCVQYSR 2759
Qy 334 CKL---G---QD---R---P---I--- 342
Db 2760 CDYINDCGGSDBAGCLFRDCNATTEFMCNRRRCIPREFICNGVDNCHDNTSDKNCPCD 2819

Qy 343 ---K---T---FQ--- 346
Db 2820 RTCQSGYTKCHNSNICIPRVYVLCDDNDCGDSDENPTYCTTHTCSSEFQCTSGRCIPQ 2879
Qy 347 ---GH---T--- 349
Db 2880 HWYCDQETDCFDASDEPASCGRHSERTCLADEFKCDGRCIPSEWICDNDNDGDSDEDK 2939
Qy 350 ---N---E---VNA---IK---W--- 357
Db 2940 RHOCQONCSDSFLCVNDRPPRRRCIPOSWVCDDVDCTGDYDENQNCNTRTTCSENEFT 2999
Qy 358 --- 357
Db 3000 CGYGLCIPKIFRCDRHNDGDSYSDRGCLYOTCQOQNTFCQNGRCISKTFVCDENDCGD 3059
Qy 358 ---D---PT---G---N---L--- 363
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLDNDSEKGCINECHDPSI 3119
Qy 364 ---LAS---CS---D---M--- 371
Db 3120 SGCDHNTDTLTSFYCSRPGYKLMGDKRTCDIDECTEMPFVCSQKCNVIGSYICKCA 3179
Qy 372 ---T---L--- 374
Db 3180 PGYLREPDGKTCRONSNIPIYLFISNRYLRNLITIDGYFYSLLLEGLDNVVALDFRVEK 3239
Qy 375 ---I---W---S--- 377
Db 3240 RLWIDTQROVIERMFLNKTNETIINHRLPAAESLAVDMVSRKLYWLDARLDGLFVSDL 3299
Qy 378 ---M---KQ---D---N---C---VHDLOQ---H---NK--- 392
Db 3300 NGHREMLAQHCVDANNTFCFONPRGLAH---POGYLYWADWGHRAIYIGRVGMDGTNKS 3357
Qy 393 ---E---I---YT---IK--- 398
Db 3358 VIISTKLEWNGITIDYTNLDLYWADAHLYIEYSLEGHHRHTVVDGALPHFAITIFE 3417
Qy 399 ---W---S---P---T---G--- 403
Db 3418 DTIYTDWTRTVKEKNGKYGNSRQTLVNTHTRPFDIHVYHPYRQPIVSNPCGTNNGCS 3477
Qy 404 ---PG---T---NN--- 408
Db 3478 HLCLIKPGKGFTCECPDDFTLQLSGSTYCMPCMSTQFLCANNEKCIPIWKKCDGQKD 3537
Qy 409 ---P---NA---N---LML---AS 417
Db 3538 CSDGSEALALCPQRCRLGQFCQSDGNCSTSPQTLCAHQNCPDGSDEDL---LCENHHCD 3596
Qy 418 ---A---S---F---DS---T---VR---L---W---DV 429
Db 3597 NEWQCANKKCIPESSWQCDTFNDCEDNDESDSHCASRTCPGQFCRANGRCIPQAMKCDV 3656
Qy 430 ---DRG---I---CI---H---T---L--- 438
Db 3657 DNDGSDHSDPEIEECMSAHLCDNFTEFSCNTYRCIPKWAVCNCGVDCCRDNSDEQCEE 3716
Qy 439 ---T---K---H---Q---E---P---V--- 445
Db 3717 RICHVPVDFRCNNHHCIPLRWQCDGNDGNDSCAPRECTESEFRVCNQCIPSRW 3776
Qy 446 ---Y---S---V---A---F 450
Db 3777 ICDHYNDGNDSDRDCERMRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836
Qy 451 SPDGRYL---AS---GS---F--- 462
Db 3837 -PDGAYCATMFECKNHVCIPIPYWKCDGDDGCGSDEBELHCLDVPNCSPNFRCDNNR 3895

```
QY 463 -----D-----K-----CV-H-I-----468
Db 3896 CIYSHEVCNGVDCGDTDETEBHCRRKPTKCTEYKCGNGHCIPHNDVNCDDADDCGD 3955
QY 469 W-----N-----Q-T-----G-----A-----L-V-----477
Db 3956 WSDLCGCKGKERTCAENICEQNTQLNEGFCISCTAGPENVFDRTSCLDINECEQFG 4015
QY 478 -----H-----SY-----R-----481
Db 4016 TCPQHCRNTRKGSYECVCADGFTSMRPGKRCAAGSSPLLPLPDNVRIRKYNLSRERS 4075
QY 482 -----G-----T-----G-----G-----I-EE-----V-C-----490
Db 4076 EYLQDEYIOAVDYMDPDIGLSVVYTVRGESRFGAKRAYIINFESGRNVLQEVVD 4135
QY 491 -----W-----N-----AA-----494
Db 4136 LKLKYMOPDGIADVGRHIYMSDVKNKRIEVAKLDGRYRKWMLISTDLDPAAIAVNP 4195
QY 495 -----G-----G-----D-K-V-----498
Db 4196 LGLMFTDWCGRKPKESAMNGEDRNILVPELGLWPTGLSIDYLNNDRIYWSDFKEDVIE 4255
QY 499 -----G-----A-SA-----S-----D-----G-----505
Db 4256 TIKYDGTDRRVIKAMNPSLDIFEDQLYWISKEGEVWKQKPGQKKEKTLVNPWL 4315
QY 506 -----SV-----C-V-----L-----D-----511
Db 4316 TVQRIFHQLRNVKSPNLCQKICSHLCLLRPGGYSCACPOGSSFIEGSTTECDAAIELPI 4375
QY 512 L-----R-----K 514
Db 4376 NLPPPCRMHGGNCYFDETLDPK 4398

RESULT 9
US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-388-9

Query Match 69.8%; Score 2553.6; DB 3; Length 5405;
Best Local Similarity 9.6%; Pred. No. 2.1e-31;
Matches 412; Conservative 80; Mismatches 17; Indels 3764; Gaps 356;

QY 1 M-----S-I-----SS-D-EVN-----ELV-----YR-----Y-----15
Db 1 MGLMSWILWAGATLLMGLTQEASVDLK-NTGRBEFLTAFLONYQLAYSKEYPRLLISS 59
QY 16 -----L-Q-----ES-----G-F-S-H-----23
Db 60 LSESPASVILSQADNTSKVTVPGESVWVNISAKAEMIGSKIQHAAVHSDVAISVQ 119
QY 24 -----SA-----F-----T 27
Db 120 ALNAKPDTAELTLRPIQALGTEYFVLTTPGTSARNVKEFAVVAGAAGASVTLKGSVT 179
QY 28 F-----G-----I-----K-----S-----HI-34
Db 180 FNGKFPAGDVLRLVTLQPNVVAQLOSSVDLSGSKVTASSPVAVLSGHSCAQKHHTCNHV 239
QY 35 -----SOS-----N-----I-----S-----39
Db 240 EQLLPTSANGTHVVPVTLASQSRDYDLAFVVASOATKLTYNHGGITGSRGLQAGDVVEFEV 299
QY 40 -----N-----GAL-V-P-P-AAL-----I-S-----51
Db 300 RPSWPLYLSANVGIOVLLFCTGAIRNEVTYDYLIPDVAAYCPAYVVKVSPGCEGVAL 359
QY 52 II-Q-K-----G-LQY-V-E-AEV-----S-I-----N-----67
Db 360 VVAQTKAISGLTIDGHAVGAKLTWEAVPGSEFSYAEVELGTADMIHTAETNLGLLTG 419
QY 68 -----D-G-T-L-----72
Db 420 LAKAIGYATAADGRTVLSVPBSCBGMQCAAGQRCQVVGKAGCVAESTAVCRAQGDPH 479
QY 73 -----FDGR-----P-IE-----79
Db 480 YTTFDGRRYDMMGTCSYTMVELCSEDDTLTLPFAFSEAKNEHRSRRVSYVGLVTVRAYSHS 539
QY 80 -SLS-----LID-----AV-----M 88
Db 540 VSLTRGEVGFVLVDNQRSLPVSLSBGRRLRVYQSGPRAVVELVGLVTVYDWDQLALS 599
QY 89 P-----D-V-----V-----Q-----93
Db 600 PARFQDVQVCLGNGYNGDPADDFLTDPGALAPDAVEFASWSKLDGDLDCEDGCQNNCPA 659
QY 94 -TR-Q-QAY-----R-----D-KLA-----Q-----Q-----106
Db 660 CTFGQAQHYEGDRLCGMLTKLDGFFAVCHDTLDPRLFECVYDLCVVGGERLSLCRGLS 719
QY 107 -----QA-----A-----A-----AAA-----114
Db 720 AYAAQACLELGISVDWRSPANCPLSCPANSRYELCGPACPTSCNGAAAPSNCSGRPCVEG 779
QY 115 -----A-A-AAS-----Q-Q-G-S-A-K-----NGEN-----T-----130
Db 780 CVCLPGFVAGSGGACVPASSCGCTFQGLQAPGQEVWADELCCORRCTCNGATHQVTCRDKQ 839
QY 131 -----ANGE-----EN-----GA-H-----T-----I-----AN-143
Db 840 SCPA-GERCSVQNLGCGYDPRFGTCQSGSDPHVSPDGRRRFDMGTCTCTYLLVSGCCQNA 898
QY 144 -----N-H-----T-----DMME-----150
Db 899 ALPAFVFLVENEHRGSGTQVSYTRAVRVEARGVKVAREYVPGQVLDVDDLYLPFQAADG 958
```

QY 151 -----V-D-G-----D-----V-----E----- 156
Db 959 QVQVFRQGRDAVVRTDGLTVTYDMNARVTAKPSSVAEALCGLGNFNGDPADDLALRG 1018
QY 157 -----I-----P-----P-----N-----K----- 161
Db 1019 GGOAANALAFGNSWQETREPGCATBFGDCPKLDSLVAQLOQSKNCGILADPKPPREC 1078
QY 162 -----AV-----VL-R-----G-----H----- 168
Db 1079 HSKLDPQGAVDVCYDRCLLPGQSGPLCDALATAYAAQAAGATVHPWRSEELCLSCPP 1138
QY 169 ----- 170
Db 1139 HSHYEACSYGCLPLSCGDLVPVGGGSECHGCVCEGDFALSGESCLPLASCGVHQGYH 1198
QY 171 -----E-----V----- 172
Db 1199 PPGQTFYPPGCDLSLCHQBGGLVSCSSCGPHEACQPSGSLGCVAVGSSSTCOASGDP 1258
QY 173 -----F-----I-----C-----AW-----N-178
Db 1259 HYTFDGRFRDFMGTCVYVLAQTCGTRPGLHRAVLQENAVANGRVSUTRVTITVQVAF 1318
QY 179 -----PVS-----D-L-----LASG-----S-G-D-----ST-----A-----RI--195
Db 1319 TLRLEQROQWKVTGVNGDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAVDLVYVYRVTV 1378
QY 196 ----- 196
Db 1379 PGNYQOMCGLGNYNGDPKDDPKPNGSQAGNANFNGSWEVVPDPSCLPPTPCPGS 1438
QY 197 ----- 198
Db 1439 EDCIPSHKCPPELEKKYQKEEFCGLLSSPTGPLSSCHKLVDPOGPLKDCIFDCLCGNLL 1498
QY 199 S-----E-----NS-----VL-----R----- 212
Db 1559 QDCAEBCQCDSDGLYNGQACVPIQQCGCYHNGAYYEPEQTVLIDNCRQOCTCHAGKMV 1618
QY 213 -----H-----C-----I-----R-----E-GGQ----- 220
Db 1619 QBEHCKPGQVQCPSPGSGILSCVTDKPCHGVTCRPOETCKEQQGGGVCLPNYEATCWLMD 1678
QY 221 -----D----- 224
Db 1679 PHVHFDGRKFDQGTCTNYVLAATTGCGVSTQGLTPFTVTTKQNRGNPAVSYYRVVTV 1738
QY 225 -----N-----KD-----VT-----SLDWN--234
Db 1739 ALGTNISIIHKDEIGKVRVNGVLTALPVSVDAGRISVTQAGKALLVADEGLQVSYDNWR 1798
QY 235 ----- 236
Db 1799 VDTLPSSTHYGAVGCLGNMNRPNNDQVFPNGTLAPSPINGGWSRAPGWDPLCWDBCR 1858
QY 237 G--T-----L-----LA--TG----- 243
Db 1859 GSCPTCPEDLEQYEGFGPLAGTGGPFTTCHAHVPESPFFKGCVLDCVCMGGDRDI 1918
QY 244 -----SY----- 245
Db 1919 LKALASYAAQAAGVVIDMRAQVCEITCPENSHYECGPPCPASCPAPLATPAV 1978
QY 246 -----D-GF-----A-R-I-----W-----T-----K-----DGN----- 257
Db 1979 CEGPCVEGQCQDAGFVLSADRCVPLNNGCCWANGTYHEAGSEFWADGTCQWCRCPGG 2038
QY 258 --L-----AS-----TL--GQHKG--PIF-ALK--W-----N----- 275

Db 2039 GSLVCTPASCGLEGEVCLLPSPQH--GCQPVSTA--EQAMGDPHVYVTLDGHRFPQGTCEY 2096
QY 276 -----KK-G--NF-----I-----LSA-----GV 286
Db 2097 LLSAPCHGPGPLGAENFTVTVANEHRSQAVSYTRSTLQIYNHSLTLSARWPKLQVDGV 2156
QY 287 -----DK-----TT----- 290
Db 2157 FVTLPFQDLSLLHAHLGADVVVTTTSGLSLAPDGSFVRLRVPAAYAGSLCGLCGNYNQ 2216
QY 291 ----- 292
Db 2217 DPADDLKAVGKPAQWQVGAQCGSCVSKPCPSCTPEQQESFGGPDACGVISATDGPL 2276
QY 293 ----- 294
Db 2277 APCHGLVPPAQYFQGGCLLDACQVQHPGGLCPAVATYVAAQQAQALREWRPDPFPQ 2336
QY 295 --AHT-----G-E-A--K-----Q----- 302
Db 2337 CPAHSHYELCGSDSCPGSCPSLSAPEGESACREGCVCDAGFVLSGDTCPVPGCGCLHDD 2396
QY 303 -----Q-F----- 304
Db 2397 RYYPQGTFTYPPGCDLSLRCREGGEVSCPSGCPHETCRPSGSLGCVAVGSSSTCOAS 2456
QY 305 --P-----F-----H-----S----- 308
Db 2457 GDPHYTFDGRFRDFMGTCVYVLAQTCGTRPGLHRAVLQENAVANGRVSUTRVTITVQV 2516
QY 309 A-----P-----ALD----- 313
Db 2517 ANFTLRLEQROQWKVTGVNGDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAVDLVYVYR 2576
QY 314 --V-----D-----W----- 316
Db 2577 VTPVGNVYQLMCGLCGNYNGDPKDDPKPNGSQAGNANFNGSWEVVPDPSCLPPTPCP 2636
QY 317 -----Q----- 317
Db 2637 PGSEGCIPSEBECPELEKKYQKEEFCGLLSSPTGPLSSCHKLVDPOGPLKDCIFDCLCG 2696
QY 318 -----SN-----N-TF-----A--S--CS--T-- 327
Db 2697 GNLSILCSNIHAYVSQAAGGHVEPWRNETFCPMECPQNSHYELCADTCSLGCALSAP 2756
QY 328 -----D-----MC--I-----H-----V-----C--K 335
Db 2757 LQCPDCAEBCQCDSDGLYNGQACVPIQQCGCYHNGAYYEPEQTVLIDNCRQOCTCHAGK 2816
QY 336 L-----CQ-----D-----R-----I----- 342
Db 2817 VVVCQHSCKPGQVQCPSPGSGILSCVTDKPCHGVTCRPOETCKEQQGGGVCLPNYEATCWL 2876
QY 343 -----K-----T-F-----Q----- 346
Db 2877 WGDPHVHFDGRKFDQGTCTNYVLAATTGCGVSTQGLTPFTVTTKQNRGNPAVSYYRVV 2936
QY 347 -----G-----HTNE-----VN-----A--I-----K----- 356
Db 2937 TAAALGTNISIIHKDEIGKVRVNGVLTALPVSVDAGRISVAQAGKALLVADFGVQVSYDW 2996
QY 357 -W--D--PT-----GNL-----LA-----S----- 366
Db 2997 NNRVDVTLPSSTHYGAVGCLGNMNRPNNDQVFPNGTLAPSPINGGWSRAPGWDPLCW 3056
QY 367 -C--S-----D----- 371
Db 3057 ECRGSCPTCPEDLEQYEGFGPLAGTGGPFTTCHAHVPESPFFKGCVLDCVCMGGGD 3116
QY 372 --TL--K-----I--W-----SM----- 378

```
Db 3117 HDILCKALASYAACAAAGVVIEDWRAQVGCITCPCNSHYEVGPPCPASCPSPAPLTT 3176
QY 379 -----K-Q-----D-----N-C-----VH-----D----- 386
Db 3177 PAVCEGBCVGCQCDAGFVLSADRCVPLNCGCWANGTYHEAGSEFWADGTCQWCRCG 3236
QY 387 -----L-----Q----- 389
Db 3237 PGGSGLVCTPASGLGVEGGLPSGHCQCPVSTAECQAWGDPHYVTLGDHRRPFGQTC 3296
QY 390 -----H-----N-----KE-----I-YT-----I-----KW----- 399
Db 3297 YLLSAPCHGPPLGAENFTVTVANERHGSQAVSYTRSTVLQIYNHSLTSARWPKLQVDG 3356
QY 400 -----S-P-----T-----G-----P-----GT-----N-N 408
Db 3357 VFVTLFQLDLSLHAHLSGADVVTTSGLSLAPGDSFVRLRVPAAYAGSLGCLGCNYN 3416
QY 409 -----PNA-NLM-----L-----A-----SAS----- 419
Db 3417 QDP-ADDLKAVGKPKAGVGVGAQCGECVSKPCSPCTPEQOESFGPDACGVISATDG 3475
QY 420 -----F-----D-----ST-V-----RLM----- 427
Db 3476 PLAPCHGLVPPAQVYFQGLLDACQVGHGCLCPAVATYVAACQAGALREWRPDPFCP 3535
QY 428 -----P-----D-----V----- 429
Db 3536 FQCPAHSHYELCGDSCFGSCPSLSAPGCSACREGCVCDAGFVLSGDTCPVPGQCCLH 3595
QY 430 -DR-----G-I-----CI----- 435
Db 3596 DDRIYPLGQTFYPGPGCDLSLRCRGEVSECPSPGPHETCRPSGSGSLGCVAVGSTTCQ 3655
QY 436 -----H-----T-----L-----TK-----H-----QE----- 443
Db 3656 ASGDPHYTTFDGRFDFMGTCVYVLAQTCGTRPGLHRFAVLQENAVMNGRVSVTRVITV 3715
QY 444 -----PV----- 446
Db 3716 QVANFTLRLSQRQKVTGVNDMKLPVVLANGQIRASQSGSDVVIETDFGLRVAYDLVY 3775
QY 447 -----SV-----A-F-S-----PD----- 453
Db 3776 VRVTPGNYQLMCLGNYNGDPKDFQKPGSQAGNANEFGNSWEVVPDSCPLPPT 3835
QY 454 -----G-----RY-----LAS-G-----S-----FD----- 463
Db 3836 CPPGSEGCIPSECPPELEKKYQKEEFCGLLSPTGPLSSCHKLVDPQGPLKDCIFDLCL 3895
QY 464 -----K-C-----V-----HI-W-N-T-----Q-----T-G-AL- 476
Db 3896 GGGNLSILCSNIHAYVSACQAAGHVBPWRNETFCPMECFQNSHYELCAUTCSLGCSALS 3955
QY 477 -----Y-----HS-----Y-----R----- 481
Db 3956 APLOCPDGAECQCDGFLYNGACVPIQQCGYHGVVYEPBQTVLIDNCRQQCCTHV 4015
QY 482 -----G-----TGIF-----E-----VC----- 490
Db 4016 KVVVCOEHSCKPGQVQCPGSGILSCVNDKDPCHGVTRCPQETCKEQGGQVCLPNYEATC 4075
QY 491 -----W-----N----- 492
Db 4076 WLKGDPHYSFDGRKDFQGTQNYVLTATTCGPGVSTOGLTFPTVTTKNQRGNPAVSVR 4135
QY 493 -----AA-G-----D-----KV-----CAS-----A-----S- 503
Db 4136 VVTVAALGTNISIHKDEIGKVRVNGVLTALPVSVADGRISVAQAGKALLVADFGLVSY 4195
QY 504 -----D-----GVC-----V-----L-----D-L- 512
Db 4196 DWNVRVDVTLPSYHGAVCGLCGNMDRNPNDQVFPNGTLAPISIPWGGSWRAPGWDPLC 4255
```

```
QY 513 -----R-----K 514
Db 4256 WDECRGSCPTCPE 4268
```

RESULT 10

```
US-09-334-220-1
; Sequence 1, Application US/09334220
; Patent No. 6323177
; GENERAL INFORMATION:
; APPLICANT: St. Jude Children's Research Hospital
; APPLICANT: Curran, Thomas
; APPLICANT: D'Arcangelo, Gabriella
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND
; TITLE OF INVENTION: THERAPIES
; FILE REFERENCE: 2427/0F704
; CURRENT FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3460
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-334-220-1
```

```
Query Match 69.6%; Score 2545.5; DB 3; Length 3460;
Best Local Similarity 11.5%; Pred. No. 2.7e-32;
Matches 393; Conservative 90; Mismatches 16; Indels 2915; Gaps 362;
```

```
QY 1 M-S-----I-----S-----S-----D-----EV----- 8
Db 1 MERSGWARTFLALLLGLATLRARAAGYPPRPFFFLCTHHGELEGDEQGEVLISLH 60
QY 9 ---N-----F---LV---VRY---LQ---BS-----GF---S-H- 23
Db 61 IAGNPTYYVPGQEHVHTISTSTFFDGLLVTLGL-YTSTVQASQSIGSSAFGFGIMSDHQ 119
QY 24 -----S-A-----P-----T-----P-----G-I-K----- 31
Db 120 FGNQFMCVVASHVSHLPTNLSTFIWIAPPAGTCGVNFMATATHRGQVIFKDALAQOLCE 179
QY 32 -----S-----H-----I-----S-----Q-SNI-----N-----CAL----- 43
Db 180 QGAPTDVTVHPLAEIHSDSIILRDDPDSYHQQLNPNIWVECNCTGEGCQCAIMHGNA 239
QY 44 V-----P---P-----A-AL-----I-----SI-----IQ----- 54
Db 240 VTFCEPYGPRELITTTGLNTTASVLQFSIGSGSCRSFSDPSIIVLYAKNNSADWIQLEK 299
QY 55 -----KG---LQY---V-BA-E-----VSIN----- 67
Db 300 IRAPSNVSTIIHILYLPEDAKGENVQFQWQENLRVGEVYEAACWALDNILI-INSARQV 358
QY 68 --ED-----GT-LP-----DG-----R----- 76
Db 359 VLEDSLDPVDTGNWLPFGATVKHSCQSDGNSIYFHGNEGSEFNFAATDVLSTEDIQE 418
QY 77 -----P-----IES-LSLI-----DA----- 86
Db 419 QWSEEFESQPTGWDVLGAVIGTECGTIESGLSVFLKDGGERKLCITPSMDTGTGNLRFYF 478
QY 87 VM-----P-----D-----D-----VW-----QT----- 94
Db 479 VMGGIDCPGNSHENDIILYAKIEGRKEHITLDTLSYSYKVPFSLVSVVINPELQTPATKF 538
QY 95 --RQ---QAY-R-----D-----K-LAQ-----O----- 106
Db 539 CLRQKHQGHNRNVAADFHVLPVLPSTWSEMIQFSINLGCCTHOPGNSVLSFSTNHG 598
QY 107 QA-----A-----AAA-----AA----- 114
```


QY 422 -S-----T-----VR-----L-----W-D-----VD-RG-----432
Db 2807 PSVFFPTKGWKRTYPLPESLVGNVRFYQYSDMQWAINDFYLPGLDRCRGGDC 2866
QY 433 -----IC-----I-HTL-T--KHQ--E--P-----V-----445
Db 2867 LREQICDPGYSGPCNVLHTLTKFLKERPDSEIKPDLWMSLEGSTCTECILAEDTA 2926
QY 446 -Y--S-----V-----A-F-----S--P-----DG-----454
Db 2927 LYFGGSVRAQVQDLRLCAKFLQYWGRISENNTSCHRPICKEGLVLLDYDGGIT 2986
QY 455 -----R-Y-L--A-----S-----SG-----460
Db 2987 WTLHENDYQKYSVRHDYILLPEDALTNTTLRWMPQFVISNGIVVGSVERAQMALDNI 3046
QY 461 -----S-----K-----C-----VHI-W-N-----T---471
Db 3047 LIGGAENPSQLVDTFDEGTSHENWSFYFNAVRTAGFCGNFSPHLYWPNKKDKTHNA 3106
QY 472 -----Q-----T-G-----A-----LV-----477
Db 3107 LSSRELIIQGYMQFKIVVGCATSCGLHSLVMLEYTKDARSDSWQVOTQCLPSSNS 3166
QY 478 -----H-S-Y-----R-----G--T-----483
Db 3167 IGCSPFQHEATTYNSVNSSWKRTIQLPDHVSSTATQFRWIOKGEETEKQSWAIDHY 3226
QY 484 -----G-----I--PE--VC-----W-N-492
Db 3227 IGACPKLCSGHGYCTGAICIDESFGDDCSVFSHDLPSYIKDNFESARVTEANWETI 3286
QY 493 -----A-A-GDKV--G-----ASAS--D-----GS--V--C 508
Db 3287 QGGVIGSGGQALAPYAHGDSLYFNGCQIROA-ATKPLDLTRASKIMFVLIGMSQTDSC 3345
QY 509 -----VL-----D-----L--R-K 514
Db 3346 NSDLSPGHAVDKAVLLQYSVNNGITWHVIAHQHPKDFTOQRVSNVPLEARMK 3399

RESULT 11

US-09-911-842A-2
; Sequence 2, Application US/09911842A
; Patent No. 6656707
; GENERAL INFORMATION:
; APPLICANT: Amgen Inc.
; TITLE OF INVENTION: C3B/C4B COMPLEMENT RECEPTOR-LIKE MOLECULES AND USES THEREOF
; FILE REFERENCE: 01017/37592
; CURRENT APPLICATION NUMBER: US/09/911,842A
; CURRENT FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: US 60/222,438
; PRIOR FILING DATE: 2000-08-01
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 3571
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-911-842A-2

Query Match 69.5%; Score 2541.9; DB 4; Length 3571;
Best Local Similarity 11.34; Pred. No. 3.8e-32;
Matches 403; Conservative 77; Mismatches 18; Indels 3061; Gaps 352;
QY 1 M-----SI-----S-----S-----5
Db 24 MSPSRNFSPLFPETAPAGCSIPAPPAPGDEAGSRVERLGOAFRRVRLLRELSELE 83
QY 6 -----D-----EYNF-----L-----V-----Y--R--Y-----15
Db 84 LVFLVDDSSSVGEVNFSELMFVKLLSDFPVPTATRAVATVTFSSKNYVVRVDYISTR 143

QY 16 -----LOE-----S-----G-F--S--H-----S-----24
Db 144 RARQHKCALLQEIIPAISYRGGTYTKGAFQAAQIILLHARENSTKVFLITDGYSNNGD 203
QY 25 -----A-----FTFGI-----K-----S-----H--33
Db 204 PRPIAASLRDSGVEIFTFGIWQGNIRELNDMASTPKEEHCYLLHSFEEFEALARRALHED 263
QY 34 -----I-----S-----Q-----36
Db 264 LPSGSIQDDMVHCSYLCDEKDCDRMGSCRGTHTGHFECEICEKYGKGLQVECTAC 323
QY 37 -----SNI--N-----G-A--A-----LV--P 45
Db 324 PSPTYKPEGSPGSISSCIPCPDENHTSPPGSTSPEDCVCREGYRAGQTCVELVHCFAALXP 383
QY 46 P-----AA-----LI--SI-----Q 54
Db 384 PENGYFIQNTCNNHFNACGVRCHPGFDLVGSSIIILCLPNGLWSGSESYCRVRTCPLHRQ 443
QY 55 -----K-----G--L-----57
Db 444 PKGHISCTREMLYKTKCLVACDEGRLEGSDKLTTCQNSOWDGPPEPRCVERHCFTOM 503
QY 58 -----Q-----YV-----60
Db 504 PKDIIISPHNCQKPAKFGTICVUSCRQFILLGVKEMLRCTTSGKWNVGVQAAVCKDVE 563
QY 61 -----EA-----E-VS-----I--66
Db 564 APQINCPKIDIEAKTLBQODSANVTWQIPTAKDNGSEKSVHVHPAFTPPVFPVGDVAIV 623
QY 67 -----N-----E--D--G--70
Db 624 YTATDLSGNQASCIHIVKIDAEPPVIDWCRSPPPVQVSEKVAASWDEPQFSDNSGAEL 683
QY 71 -----T--LF-----D-G-R-----P--I--78
Db 684 VITRSHQGLDLPQGETIVQYATDPSGNRCDIHIIVKSGCEIPFPVNGDFICTPD 743
QY 79 -----E-----S-----80
Db 744 NTGVNCTLTCLEGYDFTEGSTDKYCAVEDGVMKPTTYTTEWPCAKKRFANHGFKSPMF 803
QY 81 -----L--S-----L-----ID-----85
Db 804 YKAARCDTDLMKKFSFAFETTLGKVPSPFCSDAEDIDCRLEENLTKKYCLEYNYDYENG 863
QY 86 -AVMP-----DVVQ-----T--94
Db 864 FAIGPGWGANLDSYDDFLDTVQETATISIGNAKSSRIKRSAPLSDYKILFINITAS 923
QY 95 -----R-----QQ-----97
Db 924 VPLPDERNDTLEWENQORLLQTLTITNKLRNLKNDPMYSFOLASEILLIADNSLETCK 983
QY 98 A-----Y-----R--D-----KL-----103
Db 984 ASPFCRPGSVLRGRMCMVNCPLGTYYNLEHFTCSGRIGSYQDEEGQLECKLCPGSMYTEY 1043
QY 104 -----AQ--Q-----Q-----A--108
Db 1044 IHSRNLSDCKAQCKQGTYSYSGLETCECPGLGYQPKFGSRSCISCPENTSTVKRGAVNI 1103
QY 109 AA-----A-A--A-----A-----114
Db 1104 SACGVPCPEGKFSRSLMPCPCPRDYOPNAGKAFCLACFPYGTTFPAGSRSITECSSF 1163
QY 115 -----AAA-----AS-----QQ-----G-122
Db 1164 SSTSAABESVVPAPSLGHIKKGRHEISSQVHFECFFNCPNCHNSGTCCQLGRGYVCLCPGY 1223

123 SA-K-----NG-----EN----- 129
1224 TGLKCTDIDECSPCLNNGVCKDLVGEFICEPCSGYTGORCBENINECSSPCLNKGI 1283
130 -----T-----AN----- 133
1284 CVDGAGYRCTVKFVGLHCETEVNECQSNPCLNNAVCEQVGGFLCKCPGFLGTROG 1343
134 -----E-----ENGA-----H-----TI-----A-----N 143
1344 KNVDECLSQCKNGATCKDAGNSFRCLCAAGFTGSHCELNINECQSNPCRNQATCVDEN 1403
144 -----N-----N-----H-----T-----DM 148
1404 SYSCKQPGSGKRCETEQTGFNLDFEVSIGYVMDGLPSLHALTCTFWMKSSDDM 1463
149 -----M-----E-----V-DG----- 153
1464 NYGTPISYAVDNGSDNTLLLTIDYNGWLYVNGREKITNCPSVNDGRWHIAITWTSANGI 1523
154 -----D-----V-----E-----I----- 157
1524 WKYVIDKLSGGAGLSVGLPIPGGALVLQEQDKKGEGFSPAESFVGSISQNLWDYV 1583
158 --P-----P-----N-----KA-V-----VL-----R-G-----H----- 168
1584 LSPQOVKSLATSCPEELSKGNVLAWPDLPSGIVGKVIDSKSIFCSDCPLGGSVPHLRT 1643
169 --E-----SEV-----P-----I-----C-----AWN-----P-----V-----S 181
1644 ASBDLPKSGKNLFCDFGOLFQVGNVQYCLNOGQWTPQLPCHCERISCGVPPPLENGFHS 1703
182 -DLLASG-----GDSTARI-----WN-----L-----L-----SE----- 200
1704 DDFYA-GSTVYQCNNGYLLGDS--RMFCTDNGSNVNGVSPCLDVEDCAVGDCEHAS 1760
201 --N-----S-----T-----SG-----S-----T-----Q 208
1761 CLNVDGSIYCVPPVTGDGKNCAEPIKCKAPGNPENGSHSGEITYTGAETVFSQEGYQ 1820
209 L-V-----L-----R-H-----CI-----R-----EG----- 218
1821 LMGVTKITLESGEWNHLIPYKAVSCGKPAIPENGCIIELAFTFGSKVYTRCNKGYTLA 1880
219 GQB-----V-----P-----S-----N-KDV-----T-SL-----D----- 232
1881 G-DKESCLANSWSHSPVCEPVKCSSPENINNGKYLISGLTYLSTASYCDTGYSLOG 1939
233 -----W-----NS-----EG-TL----- 239
1940 PSIICTASGIWDRAAPPACHLVFCGBPPAIPKDAVITGNFTFRNTVYTCKEGYTLAGLD 1999
240 -----LA-----T-----G-----S-----Y----- 245
2000 TIECLADGKWSRSDQCLAVSCDEPPIVDHASPETAHLFGDIAFYCDGYSLADNSQL 2059
246 -----D-----CF----- 248
2060 LMAQKQWVPPBEGQDMPRIAHFCEKPPSVSISLESVKAKFAAGSVSVFKMGEFVLN 2119
249 --ARI-----W-----T-KDG-----NL-----A----- 259
2120 TSAKIERGQGNWPMSPMSTQICIPVRGEBPPSIWNGVAGSNYSFGAMWAYSCKNGFYIK 2179
260 --ST--LQ-----HK-----G-----P-----IF-----AL-K----- 273
2180 GEKSTCEATQWSSPIPTCHPVSCGEPKPVENGFLHTTGRIFSEVRYQCNPGYKSVG 2239
274 -----W-----N-----KKG-NP-ILS-----AG-----V-D 287
2240 SPVFVCOANRHHSESPLMCVPLDCGKPPPIQNGFMKGFENFVGSKVQFFCNEGBELVD 2299
288 -----K--TT-----II----- 292

Db 2300 SSWTCQSGKWNKSNPKCMKPAKCEPPPLENLQVLKELTTEVGVVTFSCKEGHVLQGPS 2359
QY 293 -----W-D-----A-H----- 296
Db 2360 VLKCLPSQOMNDSFPVCKIVLCTPPPLISFGVPIPSALHFGSTVKYKSCVGGFLLRGNST 2419
QY 297 T-----G-----E-AK-Q-----Q-----F-P-F----- 306
Db 2420 TLOPDGTWSSPLPECVPECPQPEIPNGIIDVQGLAYLSTALYTCCKPFELVGNNTTL 2479
QY 307 H-----S-A-----P-----AL-D-----V----- 314
Db 2480 CGENGHWLGKPTCKAIECLKPKREILNGKFSYTDLHGQTVTYSCNRGPRLEGPSALTCL 2539
QY 315 --DW-----QS-N-----N-----T-----F-----A-SC----- 325
Db 2540 ETGDMVDVADPASCNAIHCDSPQPIENGFEVADYSYGAILIYSCFPQFQVAGHAMQTCEES 2599
QY 326 --S-T-----DMC-I--HV-----C-KL-----G-Q-D----- 339
Db 2600 GWSSSIPTCMPID-CGLPHIDFGDCTKLKDDQGYFEQEDDMMEVPVYTPHPVHLGAVA 2658
QY 340 -----F-----Q-----GH-----T-----N-----E----- 351
Db 2719 LPTAPENGFLRFTETSMGSAVQYSCKPGHILVGSDLRLCLLENKRWGASPRCEAISCKKP 2778
QY 352 --V-N-AIK-----Q-----L-----KI-----WSMKOD-----NC----- 383
Db 2779 NPWNGSIKGSNYTYLTYECDPGYVLNGTERRTCQDDKNWDEDEPICIPVDCSSPPV 2838
QY 360 T-----G-----N-LL-A-S-C-----S-----D----- 369
Db 2839 SANGVRGDEYTFQKEIETCNEGFLLEGARSVCLANGSWGATPDCVPVRCATPQLA 2898
QY 370 -----D-M-----T-----L-----KI-----WSMKOD-----NC----- 383
Db 2899 NGVTEGLDYGFKBKVTFHCHGYLHGAPKLTQCSGDNW---DABIPCKPVNCGPED 2954
QY 384 -----VHD-LQ-Q-----H-----N-----KE-----I-Y 395
Db 2955 LAHGPNPFSFINGHIOYQCPFGYKLGNSRRCLNSGWSGSSPSCPLPCRSTPVEY 3014
QY 396 T-----I-----K-----WS-----P-----T-G-PGTNNPN 410
Db 3015 GTVNGTDFDCGKAARIQCFKFKLLSLSEITCEADQWSSGSPHCBHTSCGSLP--MIPN 3072
QY 411 A-----N-LMLA-S-----ASPD-----S----- 422
Db 3073 AFISSETSWKENVITYSCRSYVIOGSS-DLICTEKGVMVSQPYVCEPLSCGSPSVANA 3131
QY 423 -----T-----V-R-L-W-----DVD-----R-G-----I-----C-----I- 435
Db 3132 VATGEAHTYSEVKLECLEGYTMDTDTFTCKDGRWPPERISCSKPKPLPENTHIL 3191
QY 436 H-----T-----L-----T-----K-----H----- 441
Db 3192 VHGDGDFSVNRQSVSCAEGYTFEGVNIISVCQLDGTWEPFSDSCSPVSCGKPESEHGF 3251
QY 442 -----Q-EP-----V-----YS-VAP-----SP-----D 453
Db 3252 VVGSKYTPFSTIIYQCEPGYELEGNRERVCOENROWSGGVAICKETRCETPLEFLNGKAD 3311
QY 454 -----G-----R-Y-L-----A-S-G-----S----- 461
Db 3312 IENRTTGNVTVSCNRGYSLGEPSEAHCTENGHTWHPVLPCKPNPCVPFVIPENALLSE 3371
QY 462 --P-D-----KC-----V-H-I-----WNTOT-----G-----AL--V 477

Db 3372 KEFYVDQNSIKREGFLQGHGIIITCNPDET-TOTSACEKISCGPAHVENAIARGV 3430
Qy 478 HSYR-G--T-----G--GIF--EVC-----W-----NA- 493
Db 3431 H-YOYGMITYSCYSGYMWLEG-FLRSVLENGWTSPPICRVCRFPCCQNGGICQRPNAC 3488
Qy 494 -----AG--DK--V-----G--ASAS--D--GS-----V-----CV 509
Db 3489 SCPEGWGRLECEEPICILPCLNGRCV-APVQCDPFGWTGSRCHTAVCQSPCLNGGKCV 3547
Qy 510 -----L-----DL-RK 514
Db 3548 RPNRCHCLSSWTHGNCSSRK 3566

RESULT 12

US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kasenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15281 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
US-08-471-119A-2

Query Match 69.4%; Score 2539.8; DB 2; Length 15281;
Best Local Similarity 7.6%; Pred. No. 1.1e-28;
Matches 469; Conservative 39; Mismatches 3; Indels 5662; Gaps 374;
Qy 1 M-----S-----ISSDE-----VNF-----L-----V 12
Db 5261 MKATNELSSRYAAVLHI-SDEPLPIYKIDPEAWINFEGRSLTREALAQLVKENAESV 5319
Qy 13 -----Y-----R-----YL-----Q----- 17

Db 5320 AISNIYPYKTVVERHIVRSLDQEDANAPESMDGSDWISAVRTAQOCHTSLASDLFDIA 5379
Qy 18 E-SGF--S--HS--A-F--T--F-- 28
Db 5380 EDAGFRVEVSWARQSHQSGALDAVFHHLKPAEDSRVLKFPDTHQGRPLKSLTNQPLLP 5439
Qy 29 -----G-----G-----I--KS-- 32
Db 5440 AQSRRAELLIREGLQTLPPYMPISQITLIDRMLNANGKVDRELARRAKITQKSPVE 5499
Qy 33 -----H-----H-----IS-- 35
Db 5500 DIVPPRNSVEATVCKGFTDVLGVEVGITDNFNLGCHSLMATKLAARLGRQLNTRISVD 5559
Qy 36 -----QSN--I--N-GA-- 42
Db 5560 VFDQPVVADLAAVIQRNSAPHEPIKPADYTGVPVQSFQAQRLWFLDQLNVGATWYLMPLG 5619
Qy 43 -----L-----V--P--P-- 46
Db 5620 IRLHGSILRVDALATAISALEQRHEPLRTTFHEEDGVGVVQVQDHRPKDLRIIDLSTQPKD 5679
Qy 47 -----A-----ALI--S--II-- 53
Db 5680 AYLAVLKHEQTTLFDLATEPGMRVALIRLGESEHILLSVMHHIISDGWSVEVLFDGMHRF 5739
Qy 54 -----Q--K--GL--QY--V--E--AEV-- 64
Db 5740 YSSALRQQDPMEOILPFIQYRDFAAWQKTEQVAEHQRLDYWTEHLADSTPAELLTDL 5799
Qy 65 --SI--N-----L-----F--D-- 70
Db 5800 PRESIILSGRANELPLTIEGRHLKRAFCRVHQAATPFVILLAAALRAAHYELTCAEDATLG 5859
Qy 71 T-----L-----F--D-- 74
Db 5860 TPIANRNPENLENMIGFFVNTQCMRIAIBENDNFESILVRVRSTATSAFANQDVPFESIV 5919
Qy 75 -----G--R--P--I--E--S-- 80
Db 5920 SLLPGSRDASRNPLVQVILAVHSQQDLGKLTLEGLRDEAVDSAISTRFDFVEHFLFEHAD 5979
Qy 81 -LS-----L--I-----D-----AVMP-----DV 91
Db 5980 RLSGSVLYAKELFKLRTIESVSVFLETARRALDQPLTPLAVLPLDTGVGEIASKGLDV 6039
Qy 92 -----V--Q--TR-----QO----- 97
Db 6040 PRDYPDANIVEVFOQHVRAATPDALAVKDATSILTVAQLDQSDRLAIWLSRRHMPET 6099
Qy 98 -----AV-----R-----D--KL--A-- 104
Db 6100 LVGLAPRSCETTIAMFGIMKANLAYLPLDINSIPAARLSILSAVDGNKLVLGSGVTAP 6159
Qy 105 -Q-----Q-----Q--A--A--A-- 110
Db 6160 EQENPEVEAVGIEIILAGTGLDKTQCSNAPSPATSLAYVIFTSGSTGKPKGVVVEHRSVT 6219
Qy 111 --A-----A--A--A--A--A--A-- 118
Db 6220 RLAKPSNVISKLPOGARVAHLANIAFDASITWEIATTLNGATLVCLDYHTVLDCTRLKEV 6279
Qy 119 -----S-----Q-----Q----- 121
Db 6280 FERESITVVTMLPALLKQCVAEIPETLAHLDLLYTGDRVGGHDAMRAKSLVKIGMPSGY 6339
Qy 122 G-----S-----A-----K-----N-----GE----- 128
Db 6340 GPTENTVISTIVYEDADEMFVNGVPIGKTVNSGAYVMDRNNQQLVPSGVVGVGELVVTCDGL 6399
Qy 129 -----N-----TANGE----- 134
Db 6400 ARGYDPSLKNRFIYITVNGESIRAYRTGDRVVRPHDLQIEFFGFRMDQOVKIRGHRIE 6459

QY 135 E N GA 138
Db 6460 PGEV SALLSHNSVQDAAVVICAPADQSGAEMVAFVAARNTEDDTQBEAAVDQVQGW 6519
QY 139 H T ANN H T D 147
Db 6520 THPETAAYSEVKDIROSEVGNDFMGWTSYDGSIDKTDHMLNDTMRMILDAREPGHV 6579
QY 148 M M E V DG D 154
Db 6580 LEIGTGTGMWFMNLAQCPGLQGVGFEPKSAQFVNDAAQSPALKDGRSIVHVGTAID 6639
QY 155 INKAGPIQPRLVINSVAQYFPTPEYLFRVVEALVQIPSVRIVFGDMRTNAINRDFVAS 6699
Db 6640 INKAGPIQPRLVINSVAQYFPTPEYLFRVVEALVQIPSVRIVFGDMRTNAINRDFVAS 6699
QY 159 P 159
Db 6700 RALHTLGEKANKRLVRQMIYELBANBEELTDPAFTSLRTRLGEKIKHVEILLPTMKAT 6759
QY 160 N K AVL RG H E 169
Db 6760 NELSKRYAAVLHVRGSRQSTHQVSPNAWIDFAADGLDRQTLINLLKEHKDAGTVAIG 6819
QY 170 SE V FI C 175
Db 6820 NIPYSKTIIVERFVNSLSEDDMEGQSLDGSAAVRAVMAAQCSPSLDAMDVKEIAQEA 6879
QY 176 A W N P V SD 182
Db 6880 GYOVEVSWARQWQNGALDAIPHFPPEKAGARTLIEFPTDYEGRNVTLTNRPLNSIQS 6939
QY 183 LL A S 186
Db 6940 RRLGTQIREKQTLPPYIPMIMVLDQMPVNNNGKIDRKELVRAIVAPKPRSAATRV 6999
QY 187 G S 189
Db 7000 APRNEIEAILRDEFEDVLGTEVSLDNFDFLGHLSLWATKLAARVSRRLDAHISIKDVED 7059
QY 190 D S T A RIW NL 198
Db 7060 QPVLADLAASIORESAPHEPIQORPYTGPAEQFAGRLWFLDQLNLGATWYLMPLAIRI 7119
QY 199 198
Db 7120 RGQLRVAALSAAULFALERRHETLRTTFEESDGVGVQIVGEARNSDLRVHDVSTGDDGEVL 7179
QY 199 SE 200
Db 7180 EVLRREQTVFPDLSSPFGWRVCLVKTGEEDHVLISVMHHIYDQWSVDILRGELGFYSA 7239
QY 201 N ST 203
Db 7240 ALRGQDPLHANPLPIQYRDFAAWQREAKQVEEHQRLQVWYKQVLDSTPAELLTDLPRP 7299
QY 204 S GS T Q L VLR H 213
Db 7300 SILSGRAGSYDVTIEGVSYGALQSCFRTSRVTFVLLTVFRIAHFRLTAVDATTGTPI 7359
QY 214 C I R EG 218
Db 7360 ANNRPELETLVGCFVNTQCMRISIADDNFELVQRVNVATAAVANQDVPFERIVSAL 7419
QY 219 G Q D VP 223
Db 7420 VPGSRNTRNPLVQMFVQSVEDYDQVRLEGLSVWMPGEASTRFDMFEHLVPGQKLT 7479
QY 224 S N K D V TS 230
Db 7480 GSVLYSSDLPEQGTIQNFVDFQECRLRSVLDQPLTPISVLPFSPNAISNLSLDLLEMTS 7539

QY 231 LD W 233
Db 7540 DYPRDRTVVDLFREQAAICPDSSIADVXSSQLTYAQDLQESDRVAAMLHERHPAESLVG 7599
QY 234 N S EG 237
Db 7600 VLSPRSCETIIAYFGIMKANLAYPLDVTYAPDARLAAILDTVGEERKLLLGAGVQPGIQ 7659
QY 238 T L LA TG 243
Db 7660 IPRLSATYAEALSHATTVDVTISIPQPSATSLAYVIFTSGTGKPKGVMLIEHRGIVRLVR 7719
QY 244 S Y 245
Db 7720 DTNNVFPBSSGALPVSHFSNLAWDAATWEIYTAVLNGGTVCIDRDTMLDIAALNSTFR 7779
QY 246 D 246
Db 7780 KENVRAAFPTPAFLKOCIAETPELVANLEILHTAGDRLPDGDANLAGKTAKGGIFNVLGH 7839
QY 247 G F AR 250
Db 7840 TENTAYSTFPVVGEEFTFVNGVPGRGISNSHAYIIDRHQKLPAGVMGELIILTGDGVAR 7899
QY 251 I W T K DG 256
Db 7900 GYTDSALNKDRFVYIDINGKSTWSYRTGDKARYPRDQGLEFFGRMDQMKIRGVRIEPG 7959
QY 257 256
Db 7960 EVELTLLDHKSVLAATVVVRPPNGDPEMIAFITDAEDDVQTHKAIYKHLOGILPAYMI 8019
QY 257 N LA ST 261
Db 8020 PSHLVILDQMPVDNGKVDKRLALRAQTVQKRSTAARPPRDEVEAVLCEBYSNLEEV 8079
QY 262 LG Q HK 266
Db 8080 EVGITDGFDFLGHSLATKLAARLSQNLNTRVSVKDFDQPIADLADIIRRGSHRHP 8139
QY 267 GPI FA LKW N KKG N 279
Db 8140 IPATPYTGVEOSFAGRLWFLEQLNLGASWYLMFPAIRMRGLPQTKALAVANALVHR 8198
QY 280 F I LS AGVD K TT 290
Db 8199 HEALRTTFEDHGVGVQVIOPKSSQDLRIDLSDAVDDTAYLAALKREQTTAFDLTSEP 8257
QY 291 II 292
Db 8258 GWRVSLRLGDDDYILSIVMHHIISDGTVDVLRLQELGFYSAAIRGQEPKLSQAKSLPIQ 8317
QY 293 W D 294
Db 8318 YRDFAVWQREQIQEAKQLKYWSQOLADSTPEFLTDLPRPSILSGEADAVPMVIDGT 8377
QY 295 AH TG B 299
Db 8378 VYQLLTDFCRTHQVTSFVLLAAFRTAHYRLTGLTLDATVGTPIANRNRPELEGLIGFVN 8437
QY 300 A K QO F PF 306
Db 8438 TQCMRMAISETETEFESLVQVRLTTTTEAFANQDVPFEQIVSTLLPGSRDTSRNPVQVMF 8497
QY 307 H S 308
Db 8498 ALQSQDLGRIOLEGMTDEALETPLRLDLEHVFQEVKLSGSLLYSTDLFEVETIRG 8557
QY 309 A PAL D V 314
Db 8558 IVDVFLIILRRGLEQPKQRLMAMPITDGTIKLRDQGLLTVAKPAYPRESSVIDLFRQOVA 8617
QY 315 D W QSN N T FA SC 325

Db	8618	AAPDAI	AVDSS	TLTY	ADLD	QSNK	LHML	CQEN	MAP	TLV	AV	FAP	RSC	LTV	AV	FL	GV	8677
Qy	326																	325
Db	8678	KANLAY	PLDV	NAPAR	IEA	ILS	AVP	GKLV	LVO	AH	GP	EL	GL	TW	AD	TEL	VO	8737
Qy	326																	329
Db	8738	SGD	HEQ	HAS	GP	TATS	LAY	VN	FT	SG	TG	KPK	GW	MID	HRS	I	R	8797
Qy	330																	334
Db	8798	NVSN	LA	FD	IS	VO	EI	Y	T	ALL	NG	T	V	L	C	L	D	8857
Qy	335																	339
Db	8858	LGN	PA	I	S	R	L	S	V	N	F	V	G	D	R	L	D	8917
Qy	340																	341
Db	8918	FVNG	V	P	I	G	R	I	T	S	G	A	V	M	D	G	N	8977
Qy	342																	348
Db	8978	DGE	NI	K	A	R	T	G	D	R	V	R	P	K	D	F	E	9037
Qy	349																	356
Db	9038	VVLR	K	P	A	N	O	B	E	M	A	F	I	T	S	O	E	9097
Qy	357																	363
Db	9098	FLG	T	S	M	Y	D	G	V	D	I	P	V	N	E	K	W	9157
Qy	364																	365
Db	9158	YVGL	P	A	P	S	A	A	I	F	V	N	E	A	V	K	S	9217
Qy	366																	369
Db	9218	P	T	S	E	L	I	K	V	K	V	A	V	V	P	S	V	9277
Qy	370																	374
Db	9278	L	E	S	E	B	E	L	L	V	D	P	A	F	F	S	R	9337
Qy	375																	381
Db	9338	L	I	Q	I	D	P	T	A	W	D	P	A	T	Q	K	D	9397
Qy	382																	381
Db	9398	V	N	S	L	D	T	S	W	I	S	A	R	S	A	A	I	9457
Qy	382																	388
Db	9458	A	T	D	A	N	C	E	R	S	R	V	L	H	F	P	D	9517
Qy	389																	388
Db	9518	I	I	V	L	P	O	M	T	N	A	N	G	V	K	R	A	9577
Qy	389																	388
Db	9578	I	T	D	N	F	D	M	G	H	S	L	M	A	T	K	A	9637
Qy	389																	389
Db	9638	S	V	T	S	G	P	V	E	Q	S	A	Q	R	L	F	L	9697
Qy	390																	393
Db	9698	R	T	T	E	A	Q	K	D	I	G	V	Q	V	V	H	E	9757
Qy	394																	397
Db	9758	A	R	L	G	P	T	E	Y	I	L	S	L	V	M	H	M	9817
Qy	398																	404
Db	9818	W	Q	E	A	A	V	A	E	R	O	L	A	Y	W	E	N	9877
Qy	405																	410
Db	9878	K	P	S	K	E	R	Q	V	T	L	F	S	V	L	L	T	9937
Qy	411																	414
Db	9938	L	L	T	G	T	F	F	E	S	L	V	H	R	S	V	A	9997
Qy	415																	420
Db	9998	D	I	G	N	I	T	L	E	G	L	E	H	E	R	L	P	10057
Qy	421																	425
Db	10058	E	I	L	R	R	G	L	D	O	P	O	V	S	I	S	T	10117
Qy	426																	427
Db	10118	A	V	V	D	S	E	T	S	M	S	Y	T	S	L	D	K	10177
Qy	428																	431
Db	10178	L	P	L	D	V	N	S	P	A	R	I	O	P	I	L	S	10237
Qy	432																	431
Db	10238	M	S	R	P	S	A	T	S	L	A	V	I	F	T	S	G	10297
Qy	432																	436
Db	10298	A	S	I	W	E	I	F	T	A	L	N	G	A	L	I	C	10357
Qy	437																	436
Db	10358	E	D	L	K	V	L	I	G	G	R	L	D	A	T	A	K	10417
Qy	437																	438
Db	10418	N	A	V	S	S	G	A	I	M	D	O	K	R	L	V	P	10477
Qy	439																	444
Db	10478	R	T	G	D	R	T	R	P	K	D	S	I	E	F	F	G	10537
Qy	445																	447
Db	10538	E	T	E	M	I	G	F	V	S	M	A	S	D	R	F	S	10597
Qy	448																	452
Db	10598	N	G	N	L	D	K	A	E	M	E	E	L	D	T	M	O	10657
Qy	453																	458
Db	10658	S	A	A	L	F	V	D	K	A	A	D	P	F	G	L	T	10717
Qy	459																	464
Db	10718	S	E	L	I	H	M	T	S	V	K	T	I	F	F	G	D	10777
Qy	465																	468
Db	10778	D	P	A	F	F	T	S	L	T	S	Q	W	P	K	V	H	10837

QY 469 ---W---NT-----471
Db 10838 EENAWIDPASSGMDRHVQLMDERRDAKTAIGNIPHSNTINERHFTTSLDTGEGIAQ 10897
QY 472 ---Q---T-----GAL--V--H--478
Db 10898 DSLDGSAWOSATKAMAAARCPCLSVTELVEIGQAAGFRVEVSWARQSORHGALDVFHLE 10957
QY 479 ---S---S---Y-----480
Db 10958 DDRVGRVLNFPDTERLPSTGLTSRPLORIONRRFESQIREOLOTLPLPPYWPGRIVV 11017
QY 481 ---R-----G-T-----483
Db 11018 LERMLNANSKVDKELARKARTLOTIKPSATRVAPRNDIEAVLCDFQAVLGVTVGMD 11077
QY 484 ---CG---IP-----E-----488
Db 11078 NPFELGGHSLMATKLAARLSRLDTRVSVKDIENQFILODLADVQVOTGSAPHEAIPSTPY 11137
QY 489 ---VC---W---NA---A---G-----495
Db 11138 SGPEVQSFQGRUWFLDQLNLNASWTHMPLASRLGRLRIEALQSMALATIEARHESLRTT 11197
QY 496 ---D---K---V-G-----A---SA-----502
Db 11198 FEEQDGVVPQIVRAARNKQLRIIDVSGTEDAYLAALKQEQDAAFDLTAEPGWRVALLRG 11257
QY 503 ---SG-SV-----507
Db 11258 PDDHVLISVMHHIISDGSVDILRLQBLGOLYSNASSQAPAPLPIQYRDFAIWQKDSQIAE 11317
QY 508 ---C-----508
Db 11318 HOKQLNWKRLQVNSKPAELLADFTPRKALSGDADVIPIEDQVYQNLRSFCRAHVTSS 11377
QY 509 -V-----L---D-----LR---K 514
Db 11378 FVALLAAFRAAHYRLTGAEDATIGSPIANRNPRLSGLIGCFVNTQCLRIPVK 11430

RESULT 13

US-08-460-751-2
; Sequence 2, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Glucksmann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; CITY: New York
; STREET: 1155 Avenue of the Americas
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7638-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-460-751-2

Query Match 69.4%; Score 2539.5; DB 2; Length 4303;
Best Local Similarity 9.9%; Pred. No. 1.2e-31;
Matches 414; Conservative 81; Mismatches 14; Indels 3675; Gaps 356;

QY 1 MSIS-----S---D-----EVN---F---10
Db 72 LDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEGIFANLNLSEINLSGNPFECDC 131
QY 11 -LVY-----R-----Y-----15
Db 132 GLAWLPQWAEQOVRVVPQEAATCAGPGSLAGQPLLGILPLDSCGGEYVACLPDNSSGT 191
QY 16 -----LQ-----E-----SG-20
Db 192 VAAVFSAAHEGLLOPEACSAFCSTGQGLAALSEQGWCGLCGAASPSSASFACLSICSGP 251
QY 21 -----F-S-----H---S---AF-----T---FG---29
Db 252 PAPPAPTCKGPTLLQHVFPASPATLVGPHGLASQLAFAHIAAPLVDTDTWDGDS 311
QY 30 --I-----KSH-----I---IQ--K---G---L---57
Db 312 AEVDAAGPAASHRYVLPGRYHVTAVALGAGSALLGTQVQEAAPAALELVCPSSVQSD 371
QY 39 -----I-N---G-----AL---V---P-----P-----46
Db 372 SLDSLQNRGGSGLEAAYSIVALGEEPARAVHPLCPESDTEIFPGNGHCYRLVVEKAALQ 431
QY 47 -----AA-----LIS-----I---IQ--K---G---L---57
Db 432 AQEQOQAWAGALAMVDSPAVORFLVSRVTRSLDWIGFSTVQGVVEGPAQGEAFSLES 491
QY 58 -Q-----EDG-----YV-E-----AE---V---S-65
Db 492 CQNWLPGEHPATAEHCVRGLGTMCNTDLCSPHVSVCLOPGGPVQDAENLLVGAPSG 551
QY 66 -----IN---EDG-----TL--F-----73
Db 552 DLQGLTPLAQDGLSAPHEVVMVFPGLRLSREAFLTAEFGTQELRRPAQLQVYR 611
QY 74 -----D-----GR--P---I-----78
Db 612 LLSTAGTPENGSEPSRSPDNRTQLAPACMPGRCWPCGANICLPDASCHPOCANGCTS 671
QY 79 -----E-S-----LSLI-----DAV--M--P--DVV-----92
Db 672 GFLPGAPYALWREFLFSVPAGPPAQS-VTLHGQD-VLMLPGDLVGLQHDAGPGALLHC 729
QY 93 -----Q---T-----R---Q---Q-----97
Db 730 SPAGHPGPRAPYLSANASSWLPPLPAQLEGTWGCPACALRLAAREQLTVLLGLRPNPG 789
QY 98 -----A-----Y---RD-KL-----A--Q-105
Db 790 LRLPGRYEVAEYGVNGVSRHNLSCSFDVVSPVAGLVYIPAPRDRGLRYPTNGSALVLQV 849
QY 106 -----QQ-----A---A---A-----A-----112

QY 381 ---D-N---C---V---H---D-L---Q--- 388
Db 3068 EPTADVNYVMLTCAVCLVTVMYMAAILHLKLDQDASGRAIPCGQGRFKYVILLVKTG 3127
QY 389 --- 388
Db 3128 WGRGSGTAAHVGMILYGVDSRSHRHLDDGDFRHRNSLDIFRIATPHSLGSLVMKIRVWHD 3187
QY 389 ---QH---N---KE--- 393
Db 3188 NKGLSPAFLQHVIVRDLQARSFAFLVNDWLSVETEANGGLVEKEVLAASDAALLRPRR 3247
QY 394 ---IV---T-I-K---W--- 399
Db 3248 LLVAELQGFDFKHIMLSINDRPRSFRTRIQRATCCVLLICLFLGANAVMYGAVGDSAY 3307
QY 400 ---SPT--- 402
Db 3308 STGHVSRSLPVDVTVAAGLVSSVVPVYLAILFLFRMSRSKVAGSPSPAGQVLDI 3367
QY 403 ---G---P-GT---NNP 409
Db 3368 DSCLDSSVLDSSFLTSGLHAEQAFVGOMKSDLFLLDDSKSLVCPSPGEGTLSHPDILLSDP 3427
QY 410 ---NANI---M---LAS---A-SF--- 420
Db 3428 SIVGSNLRLARGAAGHLGPEEDGFLSAPYSPAKSFASDEDLIQVLAEGVSPAPT 3487
QY 421 ---D---ST---V--- 424
Db 3488 QDTHMETDLSLSSTPGKTTETLQRLGELPPSPGLNWEQPOAARLSRTGLVEGLRK 3547
QY 425 RL---W--- 427
Db 3548 RLLPANCASLAHGLSLLLVAVAVSVGWGASFPFGVSVAVWLSSSASFLASFLGWEPLX 3607
QY 428 ---DVD--- 431
Db 3608 VLLEALYFSLVAKRLHPDED-DTLVESPAVTPVSARVVRPPHGFALFLAKEARKVKR 3666
QY 432 ---G---I---C---I---HT---L---T--- 439
Db 3667 LHGMRLSLLVYMLFLVTLTASYGDSCHGHAYRLOSIAKQELHSRAFLAITRSEELWPW 3726
QY 440 K---H-Q---E---P---V---Y---SVA---PS--- 451
Db 3727 MAHVLLPYVHGNSSPELGPRLRQVRLQALYPPDPGPRVHTCSAAGGFSTSDYVGWE 3786
QY 452 ---PD---G--- 454
Db 3787 SPHNGSGTWAYSAPDLLGAWSGSCAVYDSGGYVQELGLSLEBRDLRLFLQLHNLNDR 3846
QY 455 ---RY---L---ASG---S---PD---K---C 465
Db 3847 SRAVLELTRYPAGVLHAAVTLREFPAAAGRALAALSVRFPALRRLSAGLSPLTTSVC 3906
QY 466 ---VH---I---W---NT---Q--- 472
Db 3907 LLLFAVHFAVAEARTWHREGRWVRLGAWARMLLVALTATALVRLAQLGAADROWTRF 3966
QY 473 ---T---G---A---LV---H---S---Y-R--- 481
Db 3967 VRGRPRFTSFDOAVHSSAARGLAASLLFLLVKAAQHVRFVROHSVFGKTLCRALPEL 4026
QY 482 -G-T-G---G---IF---E--- 489
Db 4027 LGVTGLGLVGLVAYAGLAILLVSSCDVLSWSVAQALLVLCPTGLSTLCPAESWHLSPLL 4086
QY 490 C---WNA---A-G---D--- 496
Db 4087 CVGLNAILRMGLALGAVILRWYHIALRGELYRPANEPQDYEMVELFLRLRLRMGLSKV 4146
QY 497 ---KV---G---A-S-AS---DG-SV--- 507

Db 4147 KEFRHKVRFEGMEPLFRSRSGSKVSDVPPPSAGSDASHPSTSSSQDGLSVSLGRIGT 4206
QY 508 -C---V---L---D---L---RK 514
Db 4207 RCBPEPSRLQAVPEALLTPQDLNQATVEDVYQLEQLHSLQGRR 4250
RESULT 14
US-09-479-467A-2
; Sequence 2, Application US/09479467A
; Patent No. 6723557
; GENERAL INFORMATION:
; APPLICANT: Steirberg, Paul W.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE HOMOLOGS REQUIRED FOR MALE MAT
; TITLE OF INVENTION: BEHAVIOR IN NEMATODES AND ASSAYS BASED THEREON
; FILE REFERENCE: 18021-2901B
; CURRENT APPLICATION NUMBER: US/09/479, 467A
; CURRENT FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,127
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 4303
; TYPE: PRT
; ORGANISM: Homo sapiens PKD-1 protein
US-09-479-467A-2

Query Match 69.4%; Score 2539.5; DB 4; Length 4303;
Best Local Similarity 9.9%; Pred. No. 1.2e-31;
Matches 414; Conservative 81; Mismatches 14; Indels 3675; Gaps 356;

QY 1 MSIS-----S---D-----EVN---P--- 10
Db 72 LDVSHNLLRALDVGLLANLSALAEIDISNNKISTLEEGIFANLFNLSEINLSGNPFPCDC 131
QY 11 -LVY-----R-----Y----- 15
Db 132 GLAWLPQWAEQQVVRVQPEATCAGPSLAGOPLIGPLLDGCGEYVACLPDSSGT 191
QY 16 -----LQ-----E-----SG- 20
Db 192 VAAVSFAAHEGLQPEACSAFCSTGQGLAALSEQWCLCGAAQSSASFACLSLCSGP 251
QY 21 -----P-S---H---S---AF---T---FG--- 29
Db 252 PAPPAPTCTGPTLLQHVFPASPATLVGPHGLASGLAAPHIAAPLPVTDTRWDFDGS 311
QY 30 -I-----KSH-----I---S-QSN- 38
Db 312 AEVDAAGPAASHRYVLPGRYHVTAVIALGAGSALLGTDVQVEAAPALELVCPSSVQSD 371
QY 39 -----I-N---G-----AL---V---P---P--- 46
Db 372 SLDSLSTQNRGSGLEAAYSIVALGEESPARAVHPLCPSDTEIFPGNGHCYRLVVEKAALQ 431
QY 47 -----AA-----LIS-----I---IQ-K---G---L--- 57
Db 432 AQBCQWAGAAALAMVDSPAVQRFVRSVTRSLDVWIGFSTVQGVGVEGPAQGEAFSLES 491
QY 58 -Q-----YV-E-----AE---V---S- 65
Db 492 CQNWLPGEHPHATAEHCVRGLGPTGWCNTDLCSAPHSYVCELOQGGPVQDAENLLVGAPSG 551
QY 66 -----IN---EDG-----TL---F----- 73
Db 552 DLQGLPLTPLAQDQGLSAPHEVPMVFPGLRLSREAFLTAEFGTQELRPAQLRQVYR 611
QY 74 -----D-----GR---P---I----- 78
Db 612 LLSTAGTPENGSEPSRSRPNRTQLAPACMPGGRWCPGANICLPDLASCHPOACANGCTS 671

Db 2828 LVDSNPPFPFGYISNTYVSTKVASMAFOTQAGAIPIERLASERAITVKVPNNSDWAARGH 2887
QY 358 - - - - -D- - - - -P- - - - -TG- - - - -N- - - - - - - - - - -LLA- - - - - 365
Db 2888 RSSANSANSVVQPOASVGAIVTSSNPAAGLHLQNLTYLLDGHYLSBEPPEPYLAVYLH 2947
QY 366 S- - - - -CS- - - - - - - - - - -D- - - - - - - - - - - - - - - - 370
Db 2948 SEPRNEHNCASRRIRPESLQAGADHRPYTFIFSPGSRDPAGSVHNLSSHFRWSALQVS 3007
QY 371 - - - - -M- - - - -T- - - - -L- 380
Db 3008 VGLYTSLCQYFSEBDMVWRTEGLLPLEETSFRQAVCLTRHLTAFGASLFVPPSHRVFVFP 3067
QY 381 - - - - -D- - - - -N- 388
Db 3068 EPTADVNIYMLTCAVCLVTYVMAAILHLKLDQDASGRAIPFCQGRGFKYIILVKTG 3127
QY 389 - 388
Db 3128 WGRSGTTHAVGIMLYGVDSRSGHRLDGDPAFRHNSLDIFRIATPHSLGSLVKIRVWHD 3187
QY 389 - - - - -OH- - - - - - - - - - -N- - - - - - - - - - - - - - - - 393
Db 3188 NKGLSPAWFLQHVIVRDLQTARGAFVLDVNDWLSVETEANGGLVEKEVLAASDAALLRFR 3247
QY 394 - 399
Db 3248 LLVAELQGFDFKHILSIWDRPFRSFRTRIQATCCVLLICFLGANAVYGAVGDSAY 3307
QY 400 - 402
Db 3308 STCHVSRSLPSVDTVAVGLVSSVYVYVYLAIFLFRMSRKSVAAGSPSPTPAGQVLDI 3367
QY 403 - 409
Db 3368 DSCLDSSVLDSSFLLTFSGLHAEQAFVQMKSDFLDSDSKVLCVWPGSGEGLSWPDLSDP 3427
QY 410 - - - - -NANI- - - - - - - - - - -LAS- - - - -A-SF- - - - - - - - - - - 420
Db 3428 SIVGSLRLQARGOAGHGLGPEEDGFLSPSPAKSFSASDEDLIQVLAEGVSPAPT 3487
QY 421 - - - - -D- - - - -ST- 424
Db 3488 QDTHMETDLLSSLSSTPGKTELTALQRLGELGPPSPGLNWEQQAARLSRTGLVEGLRK 3547
QY 425 RL- 427
Db 3548 RLLPAWCASLAHGLSLLNVAVAVSGVWGASFPFGVSVAMLLSSASFLASFLGWEPLK 3607
QY 428 - - - - - - - - - - -DVD- -R 431
Db 3608 VLLEALYFSLVAKRLHPDED-DTLVSPAVTPVSARVVRVPHGPFALFLAKEARKVR 3666
QY 432 - - - - -G- - - - - - - - - - -I- - - - -C- - - - - - - - - - -HT- - - - -L- - - - -T- - - - - 439
Db 3667 LHGMRLSLLVYMLFLVLTLLASVGDASCHGHAYRLQSAIKQELHSAFLAITRSEELWPW 3726
QY 440 K- - - - -H- - - - -Q- - - - -E- - - - -P- - - - -V- - - - - - - - - - -SVA- - - - -FS- - - - - 451
Db 3727 MAHVLLPYVHGNQSSPELGPRRLQVRLQALYDPDPGPRVHTCSAAGGFSTSDYDVGWE 3786
QY 452 - - - - - - - - - - -PD- -G- - - - - - - - - - - 454
Db 3787 SPHNGSCTWAYSAPDLLGAWSGSCAVYDGGYVQBELGSLSEBRDLRFLQLHNLWLNDR 3846
QY 455 - - - - -RY- - - - - - - - - - -L- - - - -ASG- - - - -S- - - - -PD- - - - - - - - - - -K- - - - -C 465
Db 3847 SRAVFLFRTYSPAVGHAHVTLRFFPAAGRALAALSVRPFALRRLSAGLSPLTITSVC 3906
QY 466 - - - - -VH- - - - - - - - - - -I- - - - -W- - - - -NT- - - - - - - - - - -Q- - - - - 472

Db 3907 LLLFAVFAVAEARTWHRGRMRVLRIGAWARWLVALTAATVALRLAQLGAADROWTRF 3966
QY 473 - - - - -T- - - - - - - - - - -G- - - - -A- - - - -LV- - - - -H- - - - -S- - - - -Y- - - - -R 481
Db 3967 VRGRPRFTSFDOVAHVSSAARGLAASLLFLLLVKAAQHVRFVRQMSVFKTKTCRALPEL 4026
QY 482 -G-T-G- - - - -G- - - - - - - - - - -IF- - - - - - - - - - -E- - - - - - - - - - -V 489
Db 4027 LGVTGLVGLGVAYQAAILLVSSCVDSLMSVAQALLVLCPTGLTGLSTLCPAESWHLSPL 4086
QY 490 C- - - - - - - - - - -WNA- - - - - - - - - - -A- - - - -G- - - - - - - - - - -D- - - - - 496
Db 4087 CVGLMALRWGALRLGAVILRWYHALRGELYRPAWEPQDYEMVELFRLRLMMGLSKV 4146
QY 497 - - - - -KV- - - - - - - - - - -G- - - - - - - - - - -A- - - - -S- - - - -AS- - - - - -DG-SV- - - - - 507
Db 4147 KEPRHKVRPEGMPEPLSRSGSKVSPDPVPPSAGSDASHPTSSSSQLDGLSVSLGRGT 4206
QY 508 -C- - - - - - - - - - -V- - - - - - - - - - -L- - - - - - - - - - -D- - - - - -L- - - - -RK 514
Db 4207 RCEPEPSRLQAVFEALLTQFDRLNQATEDVYQLEQQLHSLQGR 4250

RESULT 15

US-09-052-469-8
; Sequence 8, Application US/09052469
; Patent No. 6380360
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,469
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-09-052-469-8

Query Match 69.4%; Score 2538.6; DB 3; Length 4302;
 Best Local Similarity 9.8%; Pred. No. 1.2e-31;
 Matches 409; Conservative 87; Mismatches 13; Indels 3674; Gaps 352;

Qy	1	MSIS	-----S-D-----	FN-----F-----	10
Db	72	LDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEIGFANFLNLSINLSGNPFECDC	131		
Qy	11	LVY	-----R-----	Y-----	15
Db	132	GLAWLPRWAEQQRVVQPEAATCAGPGSLAGPLIGIPLLDGGCGEYVACLPDNSSGT	191		
Qy	16	-----IQ-----	E-----	SG-20	
Db	192	VAAVFSAAHEGLLOPEACSAFCSTGQGLAALSEQGWCWCLCGAAQSSASFACLSLCSGP	251		
Qy	21	-----F-S-----	H-S-AF-----	T-PG-----	29
Db	252	PPPPAPTCRGFTLQHVFPASPGATLVGPHGLASGQLAAFHIAELPVTATRWDFGDS	311		
Qy	30	-I-----XSH-----	-----S-QSN-38		
Db	312	AEVDAAGPAASHRYVLPGRYHTAVTALGAGSALLGTDQVEAAPAALELVCPSSVQSD	371		
Qy	39	-----I-N-G-----	AL-----V-P-----	-----46	
Db	372	SLDLSIQNRGSGLEAAYSIVALGEEPARAVHPLCPSDTEIFFNGHCVRLVVEKAALQ	431		
Qy	47	-----AA-----	LIS-----I-IQ-K-G-L-57		
Db	432	AEOCQAWAGAAALMVDSPAQVRLSVRTRSLDVLWIGFVQGVGPAQGEAFSL	491		
Qy	58	-Q-----	YV-E-----AE-----63		
Db	492	QCNWLPGEHPDATAEHCVRLLGPTGWCNTDLCSAFHSYVCELPQGPVQDAENLLVAPSG	551		
Qy	64	-----V-----S-----	I-66		
Db	552	DLQGLPTLAQDGLSAPHEVEVWVFPGLRLSREALTTAEFGTOELRPAQLRQVYR	611		
Qy	67	-----N-E-----	D-----G-T-71		
Db	612	LLSTAGTPENGSEPSRDNRTQLAPACMPGGRWCPGANICLPDASCHPQACANGCTS	671		
Qy	72	-----LF-----	D-----74		
Db	672	GPGLPGAPYALWREFLFSVAAAGPPAQSVTLHGQDVLMLPGDLVGLQHDAGPGALLHCS	731		
Qy	75	-----	G-RP-----77		
Db	732	APGHPGQAPYLSANASSWLPHLPAQLGWTWACAPACALRLLAATEQLTVLLGLRPNGLR	791		
Qy	78	-----I-E-S-----	L-I-----D-85		
Db	792	MPGRYEVRAEYGVNGVSRHNLSCSFDDVWSPVAGLRVYIPAPRGRGLYVYPTNGSALV	851		
Qy	86	-----A-----	V-MP-----DV-91		
Db	852	GANATATARNPGGVSARFENVCALVATFVPGCPWETNDTLFSVVALPWLSEGEHVVDV	911		
Qy	92	V-----Q-T-----	R-----Q-Q-A-Y-----R-D-101		
Db	912	VVENSASRANLSLRTVAEPICGLRATPSPEARVLQGVLYRSPVWEAGSDMVFRTIND	971		
Qy	102	KLAQ--Q-----QAAA--AAA--AA--AS-----Q-Q-----121			
Db	972	K-QSLTFQNVFNIYQSAAVFKLSLTASNHSVNVTYNNVTVERMNRMCGLQSVTPA	1029		
Qy	122	-----G-----SA-K-----N--GE-----TA-----131			

Db 2108 PRAHSYLRPGDYRQVNASNLVSFFVAQATVTVQVLACREPEVDVVLPLQVLMRRSQRN 2167
Qy 306 F--H--SA--P--AL--DV-- 314
Db 2168 YLEAHVDLRDCVYQTEYRWEVYRTASCQPGPARVALPGVDSPRLVLPRLALPVGH 2227
Qy 315 -----D-----W-----QS--N-- 319
Db 2228 YCFVVFVSGDTLQTSIQANVTVAPELPIIEGGSYRVMSTRDLVLDGSESDPNLE 2287
Qy 320 -----N-----TF----- 322
Db 2288 DGOPTLSFHWACVASTQREAGCALNFGPRGSSVTIPRERLAAGVEYTFSLTVWKAGR 2347
Qy 323 -----A-----SC----- 325
Db 2348 KEATNTQVTLIRGRVPIVLECVSCAKQAQVYEVSRSSVYVLEGRCLNCSGSKRGWAA 2407
Qy 326 -----S--T--D-----M-----C--I-- 331
Db 2408 RTEFNKTLVLDETTSTGSGMRLVLRGVRDGEYTFTLTVLGRSGBEGCASIRLSP 2467
Qy 332 -----H-----V-----C--334
Db 2468 NRPLPGGSCRLFLPGAVALTKVHFECTGWDHDAEDAGAPLVYALLRLRCRQGHCEFCV 2527
Qy 335 -K-----LG----- 337
Db 2528 YGSLSSYGAVLPFGFRPHEVGLVAVVQDLGAADVVALNRSIAITLPEPNSATGLTVW 2587
Qy 338 -----Q--D-----R-----P----- 341
Db 2588 LHGLTASVLPGLLRQADPQHVEYSLAVTVLNEYERALDVAAEPKHERQRAQIRKNIT 2647
Qy 342 -----I-----K--T-----FQ----- 346
Db 2648 ETLVSLRVHTVDIOQIAAALAOQMGPSRELVCRCCLKOTLHKLEAMMLILOAETTAGTV 2707
Qy 347 -----G--H-----T----- 349
Db 2708 TPTAIGSILNITGDLIHLASSDVRAPOPSGELGAESPMSVQAYNLTSALMRLMSR 2767
Qy 350 --NE----- 351
Db 2768 VLNEEPLTAGBEIQAQKESDPRSLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIF 2827
Qy 352 -V-----N-----AI--K-----W----- 357
Db 2828 LVDSNPPFPFGYISNYTVSTKVSMAFQTAQAGAIPIERLASERAITVKVPNNSDMAARGH 2887
Qy 358 -----D--P--TG--N-----LLA----- 365
Db 2888 RSSANSNSVVQPSVGAVTLDSSNPAAGLHLQNLTYLLDGHYLSPEPEPYLAVYLH 2947
Qy 366 S-----CS-----D-----D----- 370
Db 2948 SEPRPNEHCASRIRPESLOQADHRPYTFIFSPGSRDPAGSYHLNLSHFWSALQVS 3007
Qy 371 --M--T--L-----K--W--S-----Q----- 380
Db 3008 VGLYTSLCQVFSSEDMVMWRTEGLLPLEETSQPRQAVCLTRHLTAFGASLFPVPSHVRVFP 3067
Qy 381 -----D--N-----C-----V-----H--D--L-----Q----- 388
Db 3068 EPTADVNYIYMLTCAVCLVTYVMMAAILHKLQDLASGRAIPFCQGRGFKYIILVKTG 3127
Qy 389 ----- 388
Db 3128 WGRSGTTAHVIMLYGVDSRSRHRHLDGDRAPHRNSLDIFRIATPHSLGSLVWIKIRWHD 3187
Qy 389 -----QH-----N-----KE----- 393
Db 3188 NKGLSPAWFLQHVIVRDLQTSARFAFLVNDWLSVETEANGGLVEKEVLAASDAALLRFR 3247

Qy 394 -----IV-----T--I--K-----W----- 399
Db 3248 LLVAELQRFDDKHIMLSIWDPRPRSRFTIQRATCCVLLI CLFLGANAVWYGAUGDSAY 3307
Qy 400 ----- 402
Db 3308 STGHVSRLSPLSDTVAVGLVSSVVVYVYLAIFLFRMSRSKVAGSPSPTPAQQVLDI 3367
Qy 403 -----G-----P--GT-----N--P-- 409
Db 3368 DSCLDSSVLDSSPLTTFSGLHAEAFVGMKSDLFDDSKSLVCWPSGEGTLLSWPDLSDPS 3427
Qy 410 --NANL--M-----LAS--A--SP----- 420
Db 3428 IVGSNLRQLARGAOGHGLGPEEDGFSLASPYSPAKSFSASDEDLIQQVLAEGVSSPAPTQ 3487
Qy 421 -----D-----ST-----V-----R 425
Db 3488 DTHMETDLSSLSSTPGEKTEETLALQRLGELGPPSPCLANWEQQAARLSRTGLVEGLRKR 3547
Qy 426 L----- 427
Db 3548 LIPAWCASLAHGLSLLVAVAVAVGWSGVPFPGVSVAMLLSSSASFLASFLGWEPLKV 3607
Qy 428 -----DVD-----R-- 431
Db 3608 LLEALYFSLVAKRLHPDED-DTILVESPAVTPVSARVPRVPPHGFALFLAKEARKVKRL 3666
Qy 432 -G-----I-----C-----I--HT--L--T-----K 440
Db 3667 HGMLRLLVYMLFLLVTLASYGDSCHGHAYRLQSAIKQELHSRAFLAITRSEELWPM 3726
Qy 441 -----H--Q--E--P--V-----Y-----SVA--FS----- 451
Db 3727 AHVLLYFVHGNOSSPGLPRLQVRQLEALYDPDPGRVHTCSAAGGFSTSDYDVGWES 3786
Qy 452 -----PD-----G----- 454
Db 3787 PHNGSGTWASAPDLLGANWSGCAVYDSGGYQELGSLSESRDLRFLQLHNWLDNRS 3846
Qy 455 -----RY-----L--ASG--S--FD-----K--C-- 465
Db 3847 RAVFELTRYSPAVGLHAAVTLRLEFPAAAGRALAALSVRPFALRLSAGLSPLLTSVCL 3906
Qy 466 --VH-----I-----W-----NT-----Q----- 472
Db 3907 LLFAVHFAVAEARTWHRGRWRVRLGAWARWLLVALTAATALVRLAQLGAADROMTRFV 3966
Qy 473 -----T-----G--A--A--LV--H-----S--Y--R----- 481
Db 3967 RGRPRFTSDQVAHVSSAARGLAASLLFLLVKAQHVRVQWVSFGTKLCRALPELL 4026
Qy 482 G-T-G-----G-----IF-----E-----VC 490
Db 4027 GVTGLGLVGLVAVQAAILLVSSCVDSLMSVAQALLVLCPTGLTGLTLCPAESWHLSPLLC 4086
Qy 491 -----WNA-----A--G-----D----- 496
Db 4087 VGLMALRLGALRGALGAVILRWRYHALRGELYPAPQDYEMVELFLRLRLMWGLSKVK 4146
Qy 497 --KV--G-----A--S--AS-----DG--SV----- 507
Db 4147 EFRHKVRFEGMEPLPSRSSRGSKVSPDVPFPPSAGSDASHPSTSSQLDGLSVSLGRIGTR 4206
Qy 508 C-----V-----L--D-----L--RK 514
Db 4207 CEPEPSRLQAVEALLTQFDRLNQATEDVYQLBQLHSLQGR 4249

Search completed: January 3, 2005, 15:28:14
Job time : 100.667 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:26:14 ; Search time 81 Seconds
(without alignments)
2282.717 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 3659

Sequence: 1 MSISDEVNPLVRYLQESG.....GDKVGSASDGSVCVLDLAK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3355.2	91.7	577	17	US-10-723-860-836
2	3265.6	89.2	542	15	US-10-264-049-2892
3	2744	75.0	395	16	US-10-363-829-446
4	2618.9	71.6	5635	15	US-10-451-168-78
5	2614.8	71.5	5636	14	US-10-032-189-128
6	2614.8	71.5	5636	14	US-10-120-801-72
7	2614.8	71.5	5636	14	US-10-023-634-93
8	2614.8	71.5	5636	16	US-10-408-765A-1895
9	2610.5	71.3	4599	15	US-10-464-368-69
10	2610.5	71.3	4599	15	US-10-464-368-70
11	2596.6	71.0	5175	14	US-10-120-801-74
12	2596.6	71.0	5175	14	US-10-369-493-6859
13	2596.6	71.0	5175	14	US-10-369-493-6861

14	2595.3	70.9	5198	14	US-10-120-801-75	Sequence 75, Appl
15	2595.3	70.9	5198	14	US-10-369-493-6858	Sequence 6858, Ap
16	2595.3	70.9	5198	14	US-10-369-493-6860	Sequence 6860, Ap
17	2595.3	70.9	5198	15	US-10-346-863-34	Sequence 34, Appl
18	2589.2	70.8	4183	13	US-10-087-192-672	Sequence 672, Appl
19	2585.9	70.7	4599	17	US-10-479-875-4	Sequence 4, Appl
20	2585.4	70.7	26926	9	US-09-759-508B-2	Sequence 2, Appl
21	2585.4	70.7	26926	17	US-10-723-860-133	Sequence 133, Appl
22	2584.9	70.6	4636	9	US-09-835-996A-33	Sequence 33, Appl
23	2584.8	70.6	4485	16	US-10-741-601-332	Sequence 332, Appl
24	2581.5	70.6	4563	15	US-10-276-774-1723	Sequence 1723, Ap
25	2580.5	70.5	4544	15	US-10-464-368-68	Sequence 68, Appl
26	2580.5	70.5	4544	17	US-10-479-875-6	Sequence 6, Appl
27	2574.9	70.4	4660	15	US-10-464-368-74	Sequence 74, Appl
28	2572.1	70.3	5985	15	US-10-296-734-810	Sequence 810, Appl
29	2570.9	70.3	4545	15	US-10-464-368-67	Sequence 67, Appl
30	2570.9	70.3	4545	15	US-10-464-368-71	Sequence 71, Appl
31	2569.5	70.2	4545	9	US-09-873-403-2	Sequence 2, Appl
32	2569.5	70.2	4545	11	US-09-750-972-2	Sequence 2, Appl
33	2566.6	70.1	4655	16	US-10-741-601-314	Sequence 314, Appl
34	2566.6	70.1	4655	17	US-10-479-875-3	Sequence 3, Appl
35	2560.7	70.0	5376	14	US-10-028-248A-74	Sequence 74, Appl
36	2560.7	70.0	5376	15	US-10-107-782-74	Sequence 74, Appl
37	2560.1	70.0	5701	9	US-09-864-761-37319	Sequence 37319, A
38	2560.1	70.0	5701	14	US-10-029-386-32438	Sequence 32438, A
39	2559.7	70.0	5374	14	US-10-028-248A-75	Sequence 75, Appl
40	2559.7	70.0	5374	15	US-10-107-782-75	Sequence 75, Appl
41	2558.3	69.9	4495	15	US-10-138-588-20	Sequence 20, Appl
42	2554.4	69.8	4753	14	US-10-369-493-5119	Sequence 5119, Ap
43	2553.6	69.8	5405	9	US-09-922-217-1116	Sequence 1116, Ap
44	2553.6	69.8	5405	13	US-10-025-380-1116	Sequence 1116, Ap
45	2553.6	69.8	5405	17	US-10-723-860-1647	Sequence 1647, Ap

ALIGNMENTS

RESULT 1

US-10-723-860-836
; Sequence 836, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 836
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-836

Query Match	91.7%	Score 3355.2;	DB 17;	Length 577;
Best Local Similarity	86.2%	Pred. No. 1.1e-48;		
Matches	456;	Conservative	38;	Mismatches 17; Indels 18; Gaps 10;
Qy	1	MSISDEVNPLVRYLQESGFSHAFTGKHSISNSNGALVPPAALISITKGLQYV	60	
Db	52	MSITSDEVNPLVRYLQESGFSHAFTGKHSISNSNGALVPPAALISITKGLQYV	111	
Qy	61	EAEVSINEDGTLDGCRPIESLSLIDAVMPDVQTRQAVRDKLAQQ--AAAAAA--AA	117	
Db	112	EAEVSINEDGTLDGCRPIESLSLIDAVMPDVQTRQAVRDKLAQQ--AAAAAA--AA	171	
Qy	118	ASQ-----QG-S-----AKNGENTANGAHTIANNHTDM-MEVDGDVEIPPKNKAVLR	166	

```
Db 172 ATAATTTAGVSHQNPSPKREATVNGENRAHSV-NNHA-KPMEIDGEVIPSQKATVLR 229
Qy 167 GHESEVEICAWNPVSDLLASGSDSTARIWNLSNSTSGSTOLVLRHCIREGGQDVPSNK 226
Db 230 GHESEVEICAWNPVSDLLASGSDSTARIWNLSNSTSGSTOLVLRHCIREGGHVDPSNK 289
Qy 227 DVTSLDWNSBGTLLATGSYDGFARITWKDGNLSTLGOHKGPFPALKWKNKGNFILLSAGV 286
Db 290 DVTSLDWNTNGTLLATGSYDGFARITWEDGNLSTLGOHKGPFPALKWKNKGNFILLSAGV 349
Qy 287 DKTTIWDHAHTGEAKQFPFPHSAPALVDVQSN--TFASCS--TDMCIHVCKLGDORPIKTF 345
Db 350 DKTTIWDHAHTGEAKQFPFPHSAPALVDVQ--NNTTFASCS--TDMCIHVCKLGDORPVKTF 408
Qy 346 QGHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQDNCVHDLOHKNKEIYTIKWSPTGPG 405
Db 409 QGHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468
Qy 406 TNNPNANMLASAFSDSTVRLWDVDRGICHTLTKHOEPVYSVAFSPDGRYLASGSFDKC 465
Db 469 TSNPNSNIMLASAFSDSTVRLWDIERGVCTHTLTKHOEPVYSVAFSPDGRYLASGSFDKC 528
Qy 466 VHIWNTQTGALVHSYRGTGGIFEVCWNAAGDKVGASDGSVCVLDLRK 514
Db 529 VHIWNTQSGNLVHSYRGTGGIFEVCWNAAGDKVGASDGSVCVLDLRK 577
```

RESULT 2

```
US-10-264-049-2892
; Sequence 2892, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2892
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2892
```

```
Query Match 89.2%; Score 3265.6; DB 15; Length 542;
Best Local Similarity 80.7%; Pred. No. 4e-47;
Matches 442; Conservative 42; Mismatches 20; Indels 44; Gaps 15;
```

```
Qy 5 SDEVNF---LV-----YR-----YLOESGFHSAFTGIKSHISQSNNG 41
Db 1 TDC--FKESKVDLQSQQKPLRTPHTLCGFCGTGLTC-GFHSFAFTFXIESHISQSNNG 57
Qy 42 ALVPPALISITOKGLOVYEAESVINEEDGLTFGRPIESLSLDVMPDVVQVROQAYRD 101
Db 58 TLVPPALISITOKGLOVYEAESVINEEDGVFDGRPIESLSLDVMPDVVQVROQAFRE 117
Qy 102 KLAQQQ--AAAAA--AAASQ-----QG-S-----AKNGENTANGENGAHTIANNHTDM 148
Db 118 KLAQQQASAAAAAATAATAATTSAGVSHQNPSPKREATVNGENRAHSV-NNHA-K 175
Qy 149 -MEVDGDEIAPPKAVVLRGHESEVFTCAWNPVSDLLASGSDSTARIWNLSNSTSGST 207
Db 176 PMEIDGEVEIPSSKATVLRGHESEVFTCAWNPVSDLLASGSDSTARIWNLSNSTSGST 235
```

```
Qy 208 QLVLRHCIREGGQDVPSNKDVTSLDWNSBGTLLATGSYDGFARITWKDGNLSTLGOHKG 267
Db 236 QLVLRHCIREGGHVDVPSNKDVTSLDWNTNGTLLATGSYDGFARITWEDGNLSTLGOHKG 295
Qy 268 PIPALKWKNKGNFILLSAGVDKTTIWDHAHTGEAKQFPFPHSAPALVDVQSN--TFASCS 326
Db 296 PIPALKWKNKGNFILLSAGVDKTTIWDHAHTGEAKQFPFPHSAPALVDVQ--NNTTFASCS 354
Qy 327 TDMCIHVCKLGDORPIKTFQHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQDNCVHD 386
Db 355 TDMCIHVCKLGDORPVKTFQHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQEVCIH 414
Qy 387 LQOHNKEIYTIKWSPTGPGTNNPNANMLASAFSDSTVRLWDVDRGICHTLTKHOEPVY 446
Db 415 LQAHNKEIYTIKWSPTGPA--TNNPNANMLASAFSDSTVRLWDIERGVCTHTLTKHOEPVY 474
Qy 447 SVAFSPDGRYLASGSFDKCVHIWNTQTGALVHSYRGTGGIFEVCWNAAGDKVGASDGS 506
Db 475 SVAFSPDGRYLASGSFDKCVHIWNTQSGNLVHSYRGTGGIFEVCWNAAGDKVGASDGS 534
Qy 507 VCVLDLRK 514
Db 535 VCVLDLRK 542
```

RESULT 3

```
US-10-363-829-446
; Sequence 446, Application US/10363829
; Publication No. US20040142331A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;
; Altus, Christina M.; Dufour, Gerard E.;
; Chalup, Michael S.; Jackson, Jennifer L.;
; Jones, Anissa L.; Yu, Jimmy Y.;
; Wright, Rachel J.; Gietzen, Darryl;
; Li, Tommy F.; Yap, Pierre E.;
; Dahl, Christopher R.; Momiya, Monika G.;
; Bradley, Diana L.; Rohatgi, Sameer D.;
; Harris, Bernard; Roseberry Lincoln, Ann M.;
; Gerstein, Jr., Edward H.; Peralta, Careyna H.;
; David, Marie H.; Panzer, Scott R.;
; Flores, Vincent Z.; Daffo, Abel;
; Marwaha, Rakesh; Chen, Alice J.;
; Chang, Simon C.; Au, Alan P.;
; Imman, Rebekah R.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PT-1183 USN
; CURRENT APPLICATION NUMBER: US/10/363,829
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: PCT/US01/27628
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/229,751
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,749
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,750
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,747
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,748
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,583
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,517
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,610
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,597
; PRIOR FILING DATE: 2000-09-06
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: PERL Program
```

```
; SEQ ID NO 446
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: LI:021759.1.orf2.2000SEP08
; US-10-363-829-446

Query Match      75.0%; Score 2744; DB 16; Length 395;
Best Local Similarity 99.5%; Pred. No. 5.4e-39;
Matches 389; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSISDEVNVLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
   |||||
Db 5 MSISDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 64
   |||||

QY 61 EAEVSINEDTGLDGRPIESLSLIDAVMPDVOTROQAYRDKLAQOQAAAAAASQ 120
   |||||
Db 65 EAEVSINEDTGLDGRPIESLSLIDAVMPDVOTROQAYRDKLAQOQAAAAAASQ 124
   |||||

QY 121 QGSAKNGENTANGEGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWNPV 180
   |||||
Db 125 QGSAKNGENTANGEGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWNPV 184
   |||||

QY 181 SLLASGSGDSTARINWLSNSSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
   |||||
Db 185 SLLASGSGDSTARINWLSNSSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244
   |||||

QY 241 ATGSDGFAIRIWKDGNLSTLQHKGPPIALKKNNKGNFILSAGVDKTTIIWDAHTGEA 300
   |||||
Db 245 ATGSDGFAIRIWKDGNLSTLQHKGPPIALKKNNKGNFILSAGVDKTTIIWDAHTGEA 304
   |||||

QY 301 KQOPPHSAPALVDQSNNTFASCSTDCMCHVKLGQDRPIKTFQGHTEVNAIKWDPT 360
   |||||
Db 305 KQOPPHSAPALVDQSNNTFASCSTDCMCHVKLGQDRPIKTFQGHTEVNAIKWDPT 364
   |||||

QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDILQHN 391
   |||||
Db 365 GNLLASCSDDMTLKIWSMKQDNCVHDILQHN 395
   |||||

RESULT 4
US-10-451-168-78
; Sequence 78, Application US/10451168
; Publication No: US20040091969A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50039
; CURRENT APPLICATION NUMBER: US/10/451,168
; PRIOR FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: PCT/US01/49232
; PRIOR FILING DATE: 2000-12-17
; PRIOR APPLICATION NUMBER: 60/256,710
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/257,048
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/260,482
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/264,922
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/266,797
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/276,988
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/281,535
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/289,622
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 110

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 5635
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-451-168-78

Query Match      71.6%; Score 2618.9; DB 15; Length 5635;
Best Local Similarity 9.6%; Pred. No. 1.3e-29;
Matches 440; Conservative 59; Mismatches 13; Indels 4071; Gaps 372;

QY 1 M-S-----I-----SS--D-----EV----- 8
   |||
Db 1 MISWEVHTVFLPALLYSLAQDASQSEIRABEIPEGASTLAFVFDVTGSMYDDLQVVI 60
   |||
QY 9 -----NF-LV----- 12
   |||

Db 61 EGASKILETSLKRPKRPLNFALVPHDPFPIGVTTTDPKKFOYELRELYVQGGDCPE 120
   |||
QY 13 -----Y-----R---Y-----LQ----- 17
   |||

Db 121 MSIGAIIKIALEISLPGSFYIVFTDASKDYRLTHEVLQIIQOKSQVVFVLTGDCDDRTH 180
   |||
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
   |||
Db 181 IGYKVEIEIASTSSGQVFLDKKQVNEVLKWBEEAVQASKVHLLSTDLHLEQAVNTWRIPF 240
   |||
QY 27 -----T----- 31
   |||

Db 241 DPSLKEVTVSLSGSPSMIBIRNPLGKLIKKGFLHLLNIHNSAKVNVNKEPEAGMWTVK 300
   |||
QY 32 ---S--H-----I-----S---Q----- 36
   |||

Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKTVSRPVQGIPTIYVLLNTSGISTPARID 360
   |||
QY 37 -----S-----NI-----N-----G----- 41
   |||

Db 361 LLELLISGSSLKTIPVKYYPHRKPYGIWNISDFVPPNEAFFLKVTGYDKDYLFRVSS 420
   |||
QY 42 -----A-----L---VP-----P-----A 47
   |||

Db 421 VFSFSSIVDPAPKVTMPKPTGYVLPQGPVDSLLPFTLSFVRNGVTLGVQDYKLKESA 480
   |||

QY 48 -----A---L-----I-----S---IQ-----KG----- 56
   |||
Db 481 SVNLDIAKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPVLOVNNVTVTGERAVLT 540
   |||

QY 57 -----L--Q-----Y---V---E----- 61
   |||
Db 541 CLIIISAVDYNLTWQRNDRVRLAEPARIRTLANLSLELKSVPKNDAGEYHCVWSSEGGSS 600
   |||
QY 62 -A-----EVS-----NE----- 68
   |||

Db 601 AASVFLTVQEPKPVMPKNQSGFTGGSEVSMCSATGYPKPKIAMVTWDMFIVGSHRYRM 660
   |||
QY 69 --DGLTF-----D-----G-----R---P-----I-- 78
   |||
Db 661 TSDGTLFIKNAAPKDAIGYGLASNSAGTDKQNSTLYRIEAPKLMVQSELLVALGDITV 720
   |||
QY 79 -E---S-----L--S--LI-----DA-----V----- 87
   |||
Db 721 MECKTSGIPPPQVWPKGDLRLRPSFTLIIDPLGLLKIQETQDLGDAGDYTCVAINAAGR 780
   |||
QY 88 -----M-----P--DV-----VQ-----T----- 94
   |||
Db 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMPFISRPFS 840
   |||
QY 95 -----R----- 95
   |||

Db 841 VSSISQLRTGALFILNLWASDKGYICEAENQFGKIQSETTVTGTGLVAPLIGISPVAN 900
   |||
QY 96 ---QQ-----A-----Y-----R---D---KL--- 103
   |||
```

Db 901 VIEGQTLPTCTLLAGNPERRWIKNSAMLLQNPYITVRSGLHIERVQLQDGGYTC 960
 Qy 104 ---A---Q-QQ---AA--- 109
 Db 961 VASVAGTNNKTSVVVHVLFTIOHQOQILSTIEGIPVTLPCKASGNPKPSVIWSKKGEL 1020
 Qy 110 ---AAA---A-A---A-AA-A--- 118
 Db 1021 ISTSAFASAGDSLVSYPGSEGEYVCTAINTAGYAKVKVQLVYVPRVFGDORG 1080
 Qy 119 -SQ---Q---GS--- 123
 Db 1081 LSQDKPVEISVLAGEEVTLPCEVKSLPPPIITWAKETQLISPPSPRHTFLPSSGSKITET 1140
 Qy 124 ---A---K---N--- 126
 Db 1141 RTSDSMYLCVATNIAGNVQAVKLVNHPVKIQRGPKHLKVQVQGVDPICNAQGTPLP 1200
 Qy 127 ---GE-N-T---A---N-G--- 133
 Db 1201 VITWSKGSMTLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEILHVQE 1260
 Qy 134 ---E---N---GA---H--- 139
 Db 1261 PPTVEDLEPPYNTTFQERVANORIEFFCPAKGTPTIKWLHNGRELTGREPGISILEDG 1320
 Qy 140 T---IA---N---N-HT---DM--- 148
 Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKQKBEQVTVSVLLNQLTNL 1380
 Qy 149 ---M---E--- 150
 Db 1381 FCEVEGTPSPILIMYKDNQVTESTTQTVNNGKILKLFATPEDAGRYSCKAINAGTS 1440
 Qy 151 ---VD---G---DV-E---I---P---PN 160
 Db 1441 QKYFNIDVLPPTLTIGTNFNEVSNVLRDVALEQCVKGTFFPDHWFKDGKPLFLGDPN 1500
 Qy 161 ---K-A---V---VL 165
 Db 1501 VELLDRGQVLHKNARNDKRYOCTVSNAGQAKDKILTIYNPPSIKGNVTTDISVL 1560
 Qy 166 ---RG--- 168
 Db 1561 INSLIKLECETRGLPMPAITWYKDGQPIMSSQALYIDKQYLHI PRAQVSDSATYTCV 1620
 Qy 169 ---E-S---E---VFI---C-A---W--- 177
 Db 1621 ANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVVIASHLSLLECKAAGNPSPIILTLWKDGV 1680
 Qy 178 ---N--- 182
 Db 1681 PVKANDNIRIEAGKKLEIMSAQEI DRGQVCVATSVAGEKEIKYEVVDVLVPPAIEGGDE 1740
 Qy 183 ---LL---ASGS---G-D---S--- 191
 Db 1741 TSYFIVMNNLLEDCHVTGSPPTIMWLKDGQLIDERDGFKILLNGRKLVIQAQVSNT 1900
 Qy 192 ---TA---R---I-W 196
 Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSGSLSERVVVVKYKPVALQCIANGIPNPSITW 1860
 Qy 197 ---NL--- 199
 Db 1861 LKDDQPVNTAGNKLIOSSGRVLQIAKTILLEDAGRYTCVATNAAGETQOQHILHVHEPPS 1920
 Qy 200 -E---N---S---T---SGST--- 207
 Db 1921 LEDAGKMLNVLVSNPVQLECKAAGNPVPIITWYKDNRLLSGSTSTWTLNRRGQIIDIES 1980
 Qy 208 -Q---L-V---LR-H---C-I--- 215
 Db 1981 AQISDAGIYKCAINSAGATELFYSLQVHVAPISGSGNNMVVVVNNPVRLECEARGIPA 2040

Qy 216 ---R---E---G---G-Q---D--- 221
 Db 2041 PSUTWLKDGSPVSSFNGLQVLSSGRILALTSAQISDTGRYTCVAVNAAGEKORDIDLVR 2100
 Qy 222 -VPSN---KD---VT---S-L- 231
 Db 2101 YVPNIMGEBEQNVSVLISQAVELLCSQDAIPPTTLTWLKDGHPLLKPKGLSISENRSVLK 2160
 Qy 232 ---D---W-N-S---EG-TLL-A 241
 Db 2161 IEDAQVQDTRGYTCEATNVAGTEKNVNVNIPPNIGSGDELTLQTLVIEGNLISLICES 2220
 Qy 242 T---GS-Y-D---G---F---A- 249
 Db 2221 SGIPPNLIWKKGSPVLTDMSGRVILSGGROLQISIAEKSDAALYSCVASNAVGTAKK 2280
 Qy 250 ---R---I---WTQD- 256
 Db 2281 EYNLVYIRPTITNSGSHPTETIIVTRGKSISLECEVQGIPTPTVTWMDGHPILKAKGVE 2340
 Qy 257 ---NL-AS---T---L---GQH- 265
 Db 2341 ILDEGHILQKNITHVSDTRGYCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPENISVVE 2400
 Qy 266 K---G-P-I-FALK-W--- 274
 Db 2401 KNSVSLTCEASGIPPLPSITWF-KOGWPSVLSNVRILSGGRMLRLMQTTMEDAGQYTCV 2458
 Qy 275 ---N---K-K---GN- 279
 Db 2459 VRNAGEERKI FGLSVLPVPHIVGENTLEDVKYKEQSVTLTCEVTGNPVPETIWHKDGQ 2518
 Qy 280 ---F-I---L---S-AGV- 286
 Db 2519 PLODEAHHIIISGRFLQITNVQVPHTRGYTCLASSPAGHSKRSFSLNPFVSPITAGVGS 2578
 Qy 287 D---K-T---T-I---I-W--- 293
 Db 2579 DGNPEDVTILNSPTSLCEAYSYPATITWFKDGPLESNRNIRILPGGRTLIQLNAQE 2638
 Qy 294 D---A---H---TG- 300
 Db 2639 DNAGRYSCVATNEAGEMIKHYEVKVIPIPIINKDGLWPGLSPEVKIKVNNLTLECEA 2698
 Qy 301 ---K---Q---Q- 304
 Db 2699 YAIPSASLSWYKDGQPLKSDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELDF 2758
 Qy 305 ---P-F--- 307
 Db 2759 DVNIQVPPSFQKLWEIGNMLDTRNGEAKDVIINNPIISLYCETNAAPPPTLTWYKDGHP 2818
 Qy 308 -SA---P---A---LDV---D- 315
 Db 2819 TSSDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLPVPIIKGANSDL 2878
 Qy 316 ---WQ---SN- 319
 Db 2879 PEEVTVLVNKSALIECLSSGSPAPRNSWQKDGQPLEDDHHKFLSNGRILQILNTQITDI 2938
 Qy 320 ---NT---PAS-C- 326
 Db 2939 GRVYCAENTAGSACKYFNLNVHVPSPVIGPKSENITVVNNFISLTCEVSGFPPLDLSW 2998
 Qy 327 ---T---D-MCI- 331
 Db 2999 LKNEQPIKLNTNTLIVPGGRTLIQIRAKVSDGGEYTCIAINQAGESKKKFSLTVVVPPSI 3058
 Qy 332 -H---V-C---K---L-GQ- 338
 Db 3059 KHDSESLSVNVNREGTSVLECESNAVPPPVITWYKNGRMITESTHVEILADGQMLHIK 3118

QY 339 -D- -R- -P- IK- -TF- 345
Db 3119 KAESDTGVVCRAINVAGRDDKNFHLNVVPPSIEGPREVIVETISNPVTLTCDATGI 3178
QY 346 -Q- -G- -H- -T- 349
Db 3179 PPPTIAWLKNHKRIENSLEVRILSGSKQIARSQSDSGNYTCIASNMEKAQKYF 3238
QY 350 -N- E- V- NA- -IKW- D- P- -T- 360
Db 3239 LSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIOWLKDKPIASGETERIRVS 3298
QY 361 -G- -NLL- A- S- -C- S- -D- -D- -MTL- 373
Db 3299 ANGSTLNIYCALTSDTGKYTCVATNPAGBEDRIFNLNVVPTPIRGKDEAEKMLTLDVT 3358
QY 374 -K- -I- W- -SM- -KQ- D- -NCV- 384
Db 3359 SINIECRATGTPPPQINWLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVTVCASN 3418
QY 385 -H- -D- -L- -Q- - 388
Db 3419 AGVDNKHYNLQVPAPPMNSMGTETITVLKGSSTSMACITDGTAPASMAWLRDQPLGL 3478
QY 389 -Q- -H- N- -KEI 394
Db 3479 DAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKLVLEPPHINGSSEHEI 3538
QY 395 -YT- -I- K- W- 399
Db 3539 SVIVNNPLELTCIASIPAPKMTWMDGRPLPOTDQVOTLGGGEVLRISTAQVEDTGRYT 3598
QY 400 -SPTG- -P- -CT- -N- 408
Db 3599 CLASSPAGDDKEYLVRVHVPNNIAGTDEPRDITVLNRQVLTCKSDAVPPPVITWLNR 3658
QY 409 -P- -N- -ANLM- LAS- -A- -S- F- 420
Db 3659 GERLOATPRVILSGRYLQINNADIGDTANYTCVASNIAGKTREFILTVNVPNNKGG 3718
QY 421 -DSTV- -RL- W- D- V- -DR- -G- -I- 433
Db 3719 POSLVILLNKSTVLECIAEGVPTPRITWRKQAGVLAGNHARYSILENGFLHIQSAHVTD 3778
QY 434 -C- -I- -H- -TL- -T- 439
Db 3779 GRYLCMATNAAGTDRRIDLQVHVPSPSAPGPTNMTVIVNVQTLTACEATGPKPSINWR 3838
QY 440 K- H- -Q- E- -P- V- -Y- -S- 447
Db 3839 KNGHLNVQDNQNSYRLLSSGSLVILSPSVDATYECTVTNGAGDDKRTVDLTQVQVPS 3898
QY 448 VA- -F- S- -P- D- 453
Db 3899 IADEPTDFLVTHAPAVITCTASGVPPSPSIHWTKNGIRLLPRDGYRILSSGAIEILATQ 3958
QY 454 -GRY- -L- -ASG- -S- F- 462
Db 3959 LNHAGRYTCVARNAAGSAHRHVTLHVHPPVQIQPSELHVLNPNILLPCEATGTPSPF 4018
QY 463 -D- -KCV- 466
Db 4019 ITWQKGINVNTSGRHNHVLPSGGLQISRAVREDAGTYMCVAQNAGTALGKILNVQVP 4078
QY 467 -H- -I- W- 469
Db 4079 PVISPHLKEYVIAVDKPIITLSCADGLPPDITWHKDGRAIVBSIRQVRLSSGLQIAFV 4138
QY 470 -N- -T- -Q- 472
Db 4139 QPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTVNENSAQLPCVADGIPTP 4198
QY 473 -T- -G- -A- -L- VH- 478

Db 4199 AINWKDNVLLANLLGKYTAEPYGELEILENVVLEDSGYFTCVANNAAGBDTHVSLTVHV 4258
QY 479 -SV- -R- -G- -T- 483
Db 4259 LPTFTFELPGDVSUNKGBQLRSLCKATGIPLPKLTWTFNNNIIPAHFDSVNGHSELVIERV 4318
QY 484 -G- -G- -I- F- -E- V- -C- 490
Db 4319 SKEDSGTYVCTAENSUGVFKAGFVYVVPKPGDYPSNWIEPLGNAILNCEVKGDP 4378
QY 491 -WN- -A- -AGD- 496
Db 4379 PTIQNRKGVDIESHIRIQLNGSLAIYGTVNEADAGDYTCVATNEAGVVERSMSTLQS 4438
QY 497 -K- V- -G- -A- -SAS- 503
Db 4439 PPIITLEPVETVINAGGKIILNCQATGEBPQPTITWSRQSHSISWDRVRNVLSNNSLYIAD 4498
QY 504 -D- -GSV- -C- V- 509
Db 4499 AQKEDTSEFEFCVARNLMGSLVRVPVIVQVHGGFSOMSAWRACSVTCGKGIQKRSRLCNQ 4558
QY 510 -L- -DL- -R- -K 514
Db 4559 PLPANGGKPCGSDLEMRNCQNK 4581

RESULT 5
US-10-032-189-128
; Sequence 128, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zetthusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shinkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: Macdougall, John R
; APPLICANT: Smithson, Glenna
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826

; PRIOR FILING DATE: 2001-03-22
 ; PRIOR APPLICATION NUMBER: 60/279,840
 ; PRIOR FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: 60/282,981
 ; PRIOR FILING DATE: 2001-04-11
 ; PRIOR APPLICATION NUMBER: 60/283,656
 ; PRIOR FILING DATE: 2001-04-13
 ; PRIOR APPLICATION NUMBER: 60/309,247
 ; PRIOR FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: 60/311,754
 ; PRIOR FILING DATE: 2001-08-17
 ; PRIOR APPLICATION NUMBER: 60/313,331
 ; PRIOR FILING DATE: 2001-08-17
 ; NUMBER OF SEQ ID NOS: 260
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 128
 ; LENGTH: 5636
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (3003)
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (3041)
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (3367)
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.
 ; US-10-032-189-128

Query Match 71.5%; Score 2614.8; DB 14; Length 5636;
 Best Local Similarity 9.5%; Pred. NO. 1.5e-29;
 Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;
 1 M-S-----I-----SS--D-----EV-----8
 1 MISVEVHTVFLFALLYSSLAQDASPOSEIRAEFFPEGASTLAFVFDVTSVMYDDLQVVI 60
 9 -----NF-LV-----12
 61 EGAKILETSLKRPKRLFNALVPHDPEIGPVITITDPKKFOYELRELYVQGGDCPE 120
 13 -----Y-----R--Y-----LQ-----17
 121 MSGAIIKALEISLPGSFIYFTDARSQYRLTHEVLQLIQKQSQVYVFLTGDCDDRT 180
 18 -----E--S--F-----S-H-S-----A-----F 26
 181 IGYKVEELASTSGQVFLDKKQVNEVLKXVEAVQASKVHLLSTDHLEQAVNTHRIFF 240
 27 -----T-----FG-----I-----K 31
 241 DPLSKEVTSLSGSPMIEIRNPLGLIKKGFGLHLLNTHNSAKVNVNKEPEAGMWTVK 300
 32 ---S--H-----I-----S--Q-----36
 301 TSSSGRHSVITGLSTIDFRAGFSRKFTLDPKTVSRPVQGIPTYVLLNTSGISTPARID 360
 37 -----S-----NI-----N-----G-----41
 361 LLELLSISGSSLKTIPVKYYPHRKPYGIWNIISDFVPNEAFFLKVTGYDKDDYLFQEVSS 420
 42 -----A-----L-----VP-----P-----A 47
 421 VSFSSIVPDAPKVTMPKTPGYVLQPGQIPCSVDSLIPFTLSFRVNGVTLGVQYLKESA 480
 48 -----A--L-----I-----S-----IIQ-----KG-----56
 481 SVSLDIKVTLSDEGFECIAVSSAGTGRAQTFFDVSPPPIQVPPNNTVTPGERAVLT 540

QY 57 -----L-Q-----Y-----V--E-----61
 DB 541 CLIIASVDYNTWQRNDRVRLAEPARIRTLANLSLELKSVDKENDAGEVHCWVSSEGSS 600
 QY 62 -A-----EVSII-----NE-----68
 DB 601 AASVFLTVQEPKPVVMPKNQSPFTGGSEVSMCSATGYPKPKIAWTVNDMFIVGSHRYRM 660
 QY 69 --DGTLP-----D-----G-----R--P-----I--78
 DB 661 TSDGTLFIKNAAPKADAGIYGLASAGTDKQNSTLRYIEAPKLMVVQSELLVALGDITV 720
 QY 79 -E--S-----L-S-LI-----DA--V-----87
 DB 721 MECKTSGIPPPQVWPKFGLDELPSTLIIDPLGLLKIQTODLDAGDYTCVAINAAGR 780
 QY 88 -----M-----P-DV-----VQ-----T-----94
 DB 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRLLDNMFIPIRPF 840
 QY 95 -----R-----95
 DB 841 VSSISQLRTGALFILNLWASDKGTIICEAENQFGKIQSETTVTVTGLVAPLIGISPVAN 900
 QY 96 -----QQ-----A-----Y-----R--D--KL--103
 DB 901 VIEGQQLTLPCTLLAGNPIDPRWIKNSAMLLONPYITVRSDGSLHIERVOLQDGGEYTC 960
 QY 104 -----A-----Q-QQ-----AA-----109
 DB 961 VASNAGTNNKTTVVVVHVLPTIHOHQOILSTIEGIPVTLPCKASGNPKPVSIVSKKGL 1020
 QY 110 ---AAA--A-A-----A-AA--A-----118
 DB 1021 ISTSAKFAGADGSLYVVSPEGESEGYCVCTATNTAGAKRKVOLTVYVRPVFGDLRG 1080
 QY 119 -SQ-----Q-----Q-----GS-----123
 DB 1081 LSQDKPEISVLAGEBVTLPCEVKSLPPIITWAKETQLISPFSPRHTPLPSGSMKITET 1140
 QY 124 -----A-----K-----N-----126
 DB 1141 RTSDSGMLCVATNAGNVTAQVKNVHPVKIQRPKHLKVQGVQORVIPCNAOQTPLP 1200
 QY 127 -----GE--N--T--A-----N--G-----133
 DB 1201 VITWSKGSMTLVGDEHHVSNPDGTLSDQATPSDAGIYTCVATNAGTDETEITLHVQE 1260
 QY 134 -----E--N-----GA--H-----139
 DB 1261 PPTVEDELPYNTTFOERVANQRIEPPCPAKGTPKPTIKWHLNGRELITGREPGISILEEG 1320
 QY 140 T--IA-----N-----N--HT--DM-----148
 DB 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKEQSVNVLLNQLTNL 1380
 QY 149 -----M-----E-----150
 DB 1381 FCEVEGTSPILIMYKDNVQVTSSTIQVNVNGKILKLFRATPEDAGRYSCKAINAGTS 1440
 QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160
 DB 1441 QKYNFIDVLVPPTIIGTNFPKEVSVVLNEDVALEQVKGTPTFPDIHWFKDGKPLFLGDPN 1500
 QY 161 -----K-A-----V-----VL 165
 DB 1501 VELLDRGQVHLHKNARNDKGRYQCTVSNAGAKQAKDKLTIYIPFSIKGNVTTDISVL 1560
 QY 166 -----RG-----H-----H-168
 DB 1561 INSLIKLECKTRGLPMPALTWTYKDGQPIMSSSQALYIDKGOYLHPIPAQVSDSATYTCV 1620
 QY 169 -----E-S-----E-----VFI-----C--A-----W-----177

Db 1621 ANVAGTAERSFHVVDVVPVPMIEGNLATPLNKQVIAHSLTLECNAAAGNPSPILTLKDG 1680
QY 178 -N- - - - -P-V--SD- 182
Db 1681 PVKANDNFRIAGGKKLEINSAQEI DRGQVYCVATSVAGEKEIKYEVDVLPVPAIEGGDE 1740
QY 183 -LL- - - - -ASGS- - - - -D- - - - -S- 191
Db 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKDGQLIDERDGFILLNGRKLKLVIAQAQVSN 1800
QY 192 -TA- - - - -R- - - - -I-W 196
Db 1801 GLYRCMAANTAGDHKEFEVTHVPTIKSSGLSERVWVKYKVALQCIANGIPNPSITW 1860
QY 197 -NL- - - - -S- 199
Db 1861 LKDDQPNTAQNKLKIQSSGRVLQIAKTILEDAGRYTCVATNAAGETQHQHQLHVHEPPS 1920
QY 200 -E- - - - -N- - - - -T- - - - -SGST- - - - - 207
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSCSTSTWTLNRCQIIDIES 1980
QY 208 -Q- - - - -L- - - - -V- - - - -LR-H- - - - -C- - - - -I- - - - - 215
Db 1981 AQISDAGIYKVAINSAGATELFYSLQVHVAPISGSNNMVAVVNNPVRLECEARGIPA 2040
QY 216 - - - - -R- - - - -E- - - - -G- - - - -G- - - - -Q- - - - -D- - - - - 221
Db 2041 PSITWLKDGSPVSSFNGLOVLSGGRILALTSTQISDTGRYTCVAVNAAGEKORDILRV 2100
QY 222 -VPSN- - - - -KD- - - - -VT- - - - -S-L- - - - - 231
Db 2101 YVPNMGEBQNVSVLISQAVELLQSDAIPPTLTWLKDGHPLLKPKGLSISENSVLK 2160
QY 232 - - - - -D- - - - -W- - - - -N- - - - -S- - - - -EG- - - - -TLL- - - - - 240
Db 2161 IEDAQVODTGRYTCATNAGTEKKYNNVWPPNIGSGDELQTLTVEGNLSILCE 2220
QY 241 AT- - - - -GS- - - - -Y-D- - - - -G- - - - -F- - - - -A- 249
Db 2221 SSGIPPNLWKKGSPVLTDMSGRVRLISGGRQLQISIAEKSDAALYSCVASNAGTAK 2280
QY 250 - - - - -R- - - - -I- - - - -WTGDG- - - - - 256
Db 2281 KEYNLQVYRPTITNSGSHPEIIVTRGKSISLECEVOGIPPTVTWMDGHPHLLKAGV 2340
QY 257 - - - - -NL-AS- - - - -T- - - - -L- - - - -GOH- - - - - 265
Db 2341 EILDEGHILQKNHVSDTGRYVCVAVNAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K- - - - -G-PI- - - - -FALK- - - - -W- - - - - 274
Db 2401 EKNSVLTCEASGIPLPSTTF- - - - -KDGWPVLSNSVRILISGGRMLRMOTTMEDAGQYTC 2458
QY 275 - - - - -N- - - - -K-K- - - - -GN- - - - - 279
Db 2459 VVRNAGEERKIFGLSVLPVPHVIGENTLEDVKVKEQSVTLTCEVTGNPVPVITWHKDG 2518
QY 280 - - - - -F-I- - - - -L- - - - -S- - - - -AGV- - - - - 286
Db 2519 OPLQDEEAHHIISGGRFLQITNVQVHTGRYTCCLASSPAGHKRSFSLNVFVSPITAGVG 2578
QY 287 -D- - - - -K- - - - -T- - - - -I- - - - -W- - - - - 293
Db 2579 SDGNPDVTILNSPTSLVCEAYSYPATITWFKDGTPLSNRNIRILPGGRTLQILNAQ 2638
QY 294 -D- - - - -A- - - - -H- - - - -TG- - - - -E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYKVYPIPIINKGDLWGLSPKEVKIKVNNLTLECE 2698
QY 300 A- - - - -K- - - - -Q- - - - -Q- - - - - 303

Db 2699 AVAIPASLSWYKDGQPLKSDDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758
QY 304 F- - - - -P-F- - - - -H- 307
Db 2759 FDNVIOVPSFQKLWELIGNMLDTGRNGEAKDVIIINNPISLYCECTNAAPPPTLTWYKDGH 2818
QY 308 -SA- - - - -P- - - - -A- - - - -LDV- - - - -D 315
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRVTCAVNEAGEDSLQYDVRVLVPIIEGANS 2878
QY 316 - - - - -WQ- - - - -SN- - - - - 319
Db 2879 LPBEVTVLVNKGALIBELSGSPAPRNSWQOGQPLEDDHHKFLNGRILQILNTOITD 2938
QY 320 -NT- - - - -FAS- - - - -C- - - - -S 326
Db 2939 IGRYVCVAENTAGSAKYFNLVNHPVPSVIGPKSENLTVVNNFISLTCEVSGFPDPDLS 2998
QY 327 - - - - -T- - - - -D- - - - -MCI- - - - - 331
Db 2999 WLKNXQPIKLTNTLIVPGGRTLQIRAKVSDGGEVTCIAINXAGESKKKFSLTVVYVPS 3058
QY 332 -H- - - - -V- - - - -C- - - - -K- - - - -L- - - - -GO- - - - - 338
Db 3059 IKDHDSLSVVNVREGTSVLECESNAVPPPVITWYKGRMITESTHVEILADGQMLHI 3118
QY 339 -D- - - - -R- - - - -P- - - - -IK- - - - - 343
Db 3119 KKAESDTCQYVCRAINVAGRDDKNFHLNVYVPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 -T- - - - -F- - - - -Q- - - - -G- - - - -H- - - - -T- - - - - 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGSKLQIARSQSDSGNYTCIASNMEGKAKYY 3238
QY 350 - - - - -N-E-V-NA- - - - -IKM- - - - -D- - - - -P- - - - -T- - - - - 360
Db 3239 FLISQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTLIQWLKDGKPIASGETERIRV 3298
QY 361 - - - - -G- - - - -NLL-A- - - - -S- - - - -C- - - - -S- - - - -D- - - - -D- - - - -MTL- - - - - 373
Db 3299 SANGSTLIYGALTSDTGKYTCVATNPAGEEDRIFNLNVYVPTIRGNKDEAKLMTYVD 3358
QY 374 - - - - -K- - - - -I-W- - - - -SM- - - - -KO- - - - -D- - - - -NCV- - - - - 384
Db 3359 TSINIECRXTGTPPPQINMLKNGCLPLSLSHIRLLAAGQVIRVRAQVSDVAVYTCVASN 3418
QY 385 - - - - -H- - - - -D- - - - -D- - - - -L- - - - -Q- - - - - 388
Db 3419 RAGVDNKHYNLQVFAPPNDMSMGTEBITVLKSGSTSMACITDGTAPSMAWLRDGOPLG 3478
QY 389 - - - - -Q- - - - -ANLM-LAS- - - - -A- - - - -S- - - - -F- - - - - 393
Db 3479 LOAHLTVTHGMVQLLKAEETDSGKYTCIASNEAGEVSKHFIKLVLEPPHNGSEHEE 3538
QY 394 I- - - - -YT- - - - -I- - - - -K- - - - -W- - - - - 399
Db 3539 ISVIVNNPLELACIASGIPAPKMTWMDGRPLQTDQVQTLGGGEVLRISTAQVEDTGRY 3598
QY 400 - - - - -SPTG- - - - -P- - - - -GT- - - - -N- - - - - 407
Db 3599 TCLASSPAGDDKEXYLVRVHVPPNIAGTDEPRDITVLRNRQVLTLECKSDAVPPPVITWLR 3658
QY 408 N- - - - -P- - - - -N- - - - -ANLM-LAS- - - - -A- - - - -S- - - - -F- - - - - 420
Db 3659 NGERLOATPRVRLSGRYLQINNADLGDGTANTYCVASNIAAGTKTREFILTVNVPNIK 3718
QY 421 - - - - -DSTV- - - - -RL-W- - - - -D- - - - -V- - - - -DR- - - - -G- - - - -I- - - - - 433
Db 3719 GPQSVILLNKSTVLECIAGVPTPRITWRKQGANLGNHARYSILENGFLHQSAHVT 3778
QY 434 - - - - -C- - - - -I- - - - -H- - - - -TL- - - - -T- - - - - 439
Db 3779 TGRYLCMATNAAGTDRRIDLQVHVPPSIAPGPTNMTVIVNVQTLTACEATGIPKPSINW 3838

```
QY 440 -K-H-----Q-E-----P-V---Y-----446
Db 3839 RKGHLNVQDNQNSYRLSSGSLVLIISPSVDDTATYECTVTNAGAGDKRTVDLTQVPP 3898
QY 447 SVA-----F-S-----P-D-----453
Db 3999 SIADPTDFLVTKHAPAVICTAGVFPFSLHWTNGIRLLPRGDDGVRILSSGAIEILAT 3958
QY 454 ---GRY-----D-----KCV-----461
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIQPPSELHVLNNPILLPCEATGTPSP 4018
QY 462 F-----D-----KCV-----466
Db 4019 FITWQKEGINVNTSGRNVAVLPSGGLQISRAVREDAGTYMCVAQNPAQTALGRIKLVQV 4078
QY 467 ---H-----I-W-----469
Db 4079 PPVISPILKEYVIAVDKPIITLSCADGLPPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 470 ---N-----T-----Q-----472
Db 4139 VQGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGIPT 4198
QY 473 ---T-----G-----A-----L-VH 478
Db 4199 PAIWKDKNVLLANLLGKYTAEPYGELILENVVLEDSGFYTCVANNAAAGBDTHVSLTVH 4258
QY 479 ---S-----Y-----R 481
Db 4259 VLPTFTLPGDVSLNKGEOURLSCKATGIPLKLITWTFNNIIPAHFDSVNGHSELVIER 4318
QY 482 ---CT-----G-----IF-----E-----V-C-----490
Db 4319 VSKEDSTYVCTAENSQVFKATGFFVYVKEPPVFKGDPSPNWIEPLGNAILNCEVKGDP 4378
QY 491 ---WN-----A-----AGD-----496
Db 4379 TPTIQNRKRGVDIEISHIRQLNGSLAIYGTVNEDAGDYTCVATNBAGVVERMSLTQ 4438
QY 497 ---K-V-----G-----A-----SAS-----503
Db 4439 SPPIITLEPVETVINAGKKIILNCOATGEPOPITWRSRQHSISWDVRVNLSSNLYIA 4498
QY 504 ---D-----GSV-----C-V-----509
Db 4499 DAQKEDTSEPECVARNLMGSLVVRVPVIVQVHGGFSQWSAWRACSVTCGKGIQKRSRLCN 4558
QY 510 --L-----DL--R-----K 514
Db 4559 QPUPANGKPCQGSDELERNCONK 4582

RESULT 6
US-10-120-801-72
; Sequence 72, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
```

```
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 72
; LENGTH: 5636
; TYPE: PRT
; ORGANISM: human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3003)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3041)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3367)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; US-10-120-801-72
```

```
Query Match 71.5%; Score 2614.8; DB 14; Length 5636;
Best Local Similarity 9.5%; Pred. No. 1.5e-29;
Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;
```

```
QY 1 M-S-----I-----SS--D-----EV-----8
Db 1 MISWEVVHTVFLFALLYSLAQDASQSEIRABEFPPEGASTLAFVFDVTCGSMYDDLQVVI 60
QY 9 -----NF-LV-----12
Db 61 EGASKILETSLKRPKPLFNALVPPHDPDPIGPVTITTDPKFQYELRELYVQGGDCPE 120
QY 13 -----Y-----R--Y-----IQ-----17
Db 121 MSIGAIIKALEISLPGSFYVFTDARSKOYRLTHEVLQIIQKQSQVVFVLTGDCDRTH 180
QY 18 -----E-----SG--F-----S-H-S-----A-----F 26
Db 181 IGYVVEETASTSSGQVFHLDKKQVNEVLKWEAEVQASKVHLLSTDHLEQAVNTRWIPF 240
QY 27 -----T-----I-----FG-----K 31
Db 241 DPSLKEVTVSLSGSPMIETRNPLGKIKKFGELHELLNHNSAKVNVNKEPEAGMWTVK 300
QY 32 ---S--H-----I-----S--Q-----36
Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKTKTSRVPQGIPTIYVLLNTSGISTPARID 360
QY 37 -----S-----S-----NI-----N-----G-----41
Db 361 LLELLSISGSSSLKTIPIKYYPHKPYGIWNIISDFVPNEAFFLKVTGYDKDDYLFQVSS 420
QY 42 -----A-----L-----VP-----A 47
Db 421 VSPSSIVDPAPKVTMPKTPGYVLPQGOIPCSVDLSLLPFTLSFRNNGVTGLGVQYLKESA 480
```

QY 48 A-L-I-S-IIO-KG-56
Db 481 SVSLDIKAVTLSDGFEYECIAVSSAGTGRAQTFFDVSEPPVQVPPNNVTVTGERAVLT 540
QY 57 L-Q-Y-V-E-61
Db 541 CLIIASVDYNTWQRDRVRLAEPARITLANLSLELKSVEKFNDAEGYHCVSSEGSS 600
QY 62 A-EVSI-NE-68
Db 601 AASVFLTVQBPVKVTPWPKQSTFGGSEVIMSATGYPKPKIAWTVMDFIVGSHRYM 660
QY 69 D-DGTLF-G-R-P-I-78
Db 661 TSDGTLFIKNAAPKDAIGYCLASNSAGTDKQNSTLRYIEAPKLMVVOSELLVALGDIIV 720
QY 79 E-S-L-S-LI-DA-V-87
Db 721 MECKTSGIPPPQVVKFKGDLRLPSTFLIIDPLLLGLLKIOETQDLDAQDYTCVAINAAGR 780
QY 88 M-P-DV-VQ-T-94
Db 781 ATGKITLDVGSPPVFIQEPADVMEIGSNVTLPFCYQGYPEPTIKWRRLDNMPFISRPFS 840
QY 95 R-95
Db 841 VSSISQRLTGALFILNLWASDKTYICEAENQFGKIQSETTVTVTGIVAPLIGISPSVAN 900
QY 96 Q-Q-A-Y-R-D-KL-103
Db 901 VIEGQOLTPLCTLLAGNPIPERWRKNSAMLLQNPYITVRSDGSLHIERVOLQDGEYTC 960
QY 104 A-Q-QO-AA-109
Db 961 VASNAGTNNKTSVVVHVLPITQHQOILLSTIEGIPVTLPCASGNPKPSVTWSKKGEL 1020
QY 110 AAA-A-A-AA-A-118
Db 1021 ISTSSAKFSAGADSLVYVSPGEESEYVCTATNTAGYAKRKVQLTIVYVRPVFGDLRG 1080
QY 119 SQ-Q-GS-123
Db 1081 LSQDKPEISVLAGEVTLPCVKSLPPPIITWAKETQLISPPSRHTFLPSSGMKITET 1140
QY 124 A-K-N-126
Db 1141 RTSDSGMYLCVATNIAGNVTAVKLVNVHVPKIQRGPKHLKVQVQORVDIPCNAOQTPLP 1200
QY 127 GE-N-T-A-N-G-133
Db 1201 VITWSKGGSTMLVDGHHVSNPDGTLSDOATPSDAGIYTCVATNIAGTDEITLHVQB 1260
QY 134 E-N-GA-H-139
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEPPCPAKGPKPTIKWLHNGRELTCREPGISILEG 1320
QY 140 T-IA-N-N-HT-DM-148
Db 1321 TLLVIAVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVFIKDKQEVNSVLLNQLTNL 1380
QY 149 M-E-150
Db 1381 FCEVEGTPSPIIMWYKDNQVVBESSTIQTVNNKILKLFRAPEADAGRYSCKAINAGTS 1440
QY 151 VD-G-DV-E-I-P-PN-160
Db 1441 OKYFNIDVLVPPTIIGNFPKEVSVLNRDVALEQVKGTPTTPIHWFKDGKPLFLGDEN 1500
QY 161 K-A-V-VL-165
Db 1501 VELLDRGOVLHLKNARRNDKGRYQCTVSNAAAGQAKDIKLTIVIPPSIKGGNVTTDISVL 1560

QY 166 RG-H-168
Db 1561 INSILKLECKTRGLPMPAITWYKDGQPIKSSSQALYIDKQYLHIAPRAQVSDSATYCHV 1620
QY 169 E-S-E-VFI-C-A-W-177
Db 1621 ANVAGTAESKSFHVDVYVPPMIEGNLATPLNKQVIAHSLTLECAAGNPSPIITLWKDGV 1680
QY 178 N-P-V-SD-182
Db 1681 PVKANDNFRIEAGGKKLEIMSARQEIQRQYICVATSVAGEKEIKYEVDVLVPPAIEGDE 1740
QY 183 LL-ASGS-G-D-S-191
Db 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKDGQLIDERDQKILLNKRKLVIQAQVSNT 1800
QY 192 TA-R-I-W-196
Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSSGLSERVVVYKYPVALQCIANGIPNSITW 1860
QY 197 NL-S-199
Db 1861 LKDDQPVTAQGNLKIQSSGRVLOIAKLTLEDAGRYTCVATNAAGETOQHIQLHVHEPPS 1920
QY 200 E-N-S-T-SGST-207
Db 1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLLSGSTSMTFLNRQIIDIES 1980
QY 208 Q-L-V-LR-H-C-I-215
Db 1981 AQTSAGIKYKCAINSAGATFYSLQVHVAPSISGSNNMVAVVNNVPVRLECEARGIPA 2040
QY 216 R-E-G-G-Q-D-221
Db 2041 PSLTWLKGSPVSSFSFNGLOVLSGGRILALTSTQISDTRYTCVAVNAAGEKORDILRV 2100
QY 222 VPSN-KD-VT-S-L-231
Db 2101 YVPPNIMGEEQNVSVLISOAVELLQSDAI PPPTLTWLDGHPHLLKPKGLSISENSVLK 2160
QY 232 D-W-N-S-EG-TLL-240
Db 2161 IEDAQVQDTRYTCBATNVAGTEKKNVNNVWPPNIGSGDELTQLTVEGNLISLCE 2220
QY 241 AT-QS-Y-D-G-F-A-249
Db 2221 SSGIPPPNLIWKKSGSPVLTDMSGRVRLSGGRQLOISIAEKSDAALYSCVASNAGTAK 2280
QY 250 R-I-WTKDG-256
Db 2281 KEYNLOVYIRPTITNSGSHPTLIIIVTRGKSI SLECEVOQIPPTVTWMDKGHPHLLKAKGV 2340
QY 257 NL-AS-L-QH-265
Db 2341 EILDEGHILQLKNIHVSDTRYGVCAVNAVAGMTDKKYDLSVHAPPSIIGNHRSPIENVV 2400
QY 266 K-G-PI-FALK-W-274
Db 2401 EKNSVSLTCEASGIPLPSTWF-KDGPVSVLSNVRIILSGGRMLRMQTTWEDAGQYTC 2458
QY 275 N-K-K-GN-279
Db 2459 VVRNAAAGEBKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPVITWHDG 2518
QY 280 F-I-L-S-AGV-286
Db 2519 QPLQDEAHIIISGGRFLQITNVQVPHTRGYTCCLASSPAGHKSRSPSLNVFVSPTIAGVG 2578
QY 287 D-K-T-T-I-W-293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLSNRNIIRILPGRTLQILNAQ 2638
QY 294 D-A-H-TG-E-299

Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGFLSPKEVKIKVNNLTLECK 2698
QY 300 A-----K-----Q-----Q----- 303
Db 2699 AVAIPASLSWYKDGOLKSDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELD 2758
QY 304 F-----P-F-----H----- 307
Db 2759 FDNIOVPPSFQKLWEIGNMLDTRNGEAKVDIINNPIISLYCETNAAPPTTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQDYRVLVLPPIEGANS 2878
QY 316 -----WQ-----SN----- 319
Db 2879 LPBEVTLVNKSALIECLSGSPAPRNSWQKQOQLLEDHFKFLSNGRILQILNTQITD 2938
QY 320 --NT-----PAS--C-----S 326
Db 2939 IGRYVCAENTAGSAKKYFNLVHVPPSVIGPKSENLTVVNNFISLTCEVSGFPFPPDL 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKXQPIKLTNTLIVPGGRTLIIRAKVSDGGEYTCIAXNAGESKKKFSLTVVVPPS 3058
QY 332 --H-----V--C-----K-----L--GQ----- 338
Db 3059 IKDHSLSLVNVNREGTSVLECESNAVPPVITWYKNGRMITESTHVEILADGQMLHI 3118
QY 339 -----D-----R-----P-1K----- 343
Db 3119 KKAESDGTGYVCRAINAGRDKNFNLVNVPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 --T-----F-----Q-----G-----H-----T----- 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGSKLOIARSOHSDSGNYTCIASNMEGKAQKY 3238
QY 350 -----N-B-V-NA-----IKW--D--P-----T----- 360
Db 3239 FLSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIQLWKDGKPIASGETERIRV 3298
QY 361 --G--NL--A--S-----C--S-----D-----D-----MTL-- 373
Db 3299 SANGSTLNIYGALTSDTGKTCVATNPAGEEDRIENLNVVTVTIRGNKDEAKLMTYVD 3358
QY 374 --K-----I-W-----SM-----KQ--D-----NCV-- 384
Db 3359 TSINIECRXTGTPPPQINLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418
QY 385 -----H-----D-----L-----Q----- 388
Db 3419 RAGVDNKHYNLQVAFPPNMDNSMGTEBITVLKSGSTSMACITDGTAPSNMWRDGOPLG 3478
QY 389 -----Q-----H-N-----KE 393
Db 3479 LDAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFILKVLPPHINGSBEHEE 3538
QY 394 I-----YT-----I--K--W----- 399
Db 3539 ISVIVNNPLETCTIASGIPAKMTWMDGRPLQTDQVQVLGGGEVLRISTAQVEDTGRY 3598
QY 400 -----SPTG-----P--GT-----N----- 407
Db 3599 TCLASSPAGDDKEYLVRVHVHPNIACTDPRDITVLNRNQVLTLECKSDAVPPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420
Db 3659 NGERLOATPRVILSGRYQIINNADLGDYANTCVASNIAGKTTREFTILTAVNPPNKG 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433

Db 3719 GPQSLVILINKSTVLECIAEGVTPTRITWRKDGAVLGNHARYSILENGFLHIQSAHVTD 3778
QY 434 -----C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHVPPIAPGPTMNTVIVNVQTTLACEATGIPKPSINW 3838
QY 440 -K--H-----Q-E-----P-V-----Y----- 446
Db 3839 RONGHLLNVQONQSVRLSSGLVLIISPSVDDTATYEYCTVINGAGDDKRTVDLTQVQPP 3898
QY 447 SVA-----F-S-----P--D----- 453
Db 3899 STADEPTFLVTKHAPAVITCTASGVPPFPIHWTNKGIRLLPRGQGYRILSSGAIFILAT 3958
QY 454 -----GRY-----L-----ASG--S-- 461
Db 3959 QLNHAGRYTCVARNAGSAHRHTLVHPPVPIQOPSELHVLNNPILLPCEATGTPSP 4018
QY 462 F-----D-----KCV----- 466
Db 4019 FITWKEGINVNTSGRHHAVLPSSGGLQISRAVREDAGTYMCVAQNPAGTALGKIKLVQV 4078
QY 467 -----H-----I-W----- 469
Db 4079 PVISPHLKEYVIAVDKPIITLSCADGLPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 470 -----N-----T-----Q----- 472
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTVNENSQAILPCVADGIFT 4198
QY 473 -----T-----G-----A-----L-VH 478
Db 4199 PAIWKKNVLLANLLGKYTAEPYGBELILENVLEDSGFYTCVANNAAGEDTHVSLTVH 4258
QY 479 -----S-----Y-----R 481
Db 4259 VLPTFTPLPGDVLNKGELRLSKATGIPKLTWTFNNIIPAHFDSVNGHSELVIER 4318
QY 482 -----GT-----G-----G-----IF-----E--V--C----- 490
Db 4319 VSKEDSGTYVTAENSVGPFVKATGTVVYKEPPVFKGDYPSNWIEPLGNAILNCEVKGDP 4378
QY 491 -----WN-----A-----AGD----- 496
Db 4379 TPTIQNRKGVDEIESHRIQLNGSLAIYGTWNEDAGDYTCVATNEAGVVERMSLTLQ 4438
QY 497 -----K-V--G-----A-----SAS----- 503
Db 4439 SPPIITLEPVETVINAGGKIILNCQATGPPQPTITWSROGHSISWDDRVNVLNNSLYIA 4498
QY 504 -----D-----GSV-----C-V----- 509
Db 4499 DAQKEDTSBFCVARNLMGSLVVRVPIVQVHGFGFSQSAWRACSVTCGKGIQKRSCLN 4558
QY 510 --L-----DL--R--K 514
Db 4559 QPLPANGKPKCGSDLEMCNQNK 4582

RESULT 7

US-10-023-634-93

; Sequence 93, Application US/10023634

; Publication No. US20030236389A1

; GENERAL INFORMATION:

; APPLICANT: Shimkets, Richard A

; APPLICANT: Colman, Steven D

; APPLICANT: Spytek, Kimberly A

; APPLICANT: Ballinger, Robert A

; APPLICANT: Guo, Xiaojia

; APPLICANT: Tchernev, Velizar T

; APPLICANT: Shenoy, Suresh G

; APPLICANT: Li, Li

; APPLICANT: Ellerman, Karen

APPLICANT: Zerhusen, Bryan D
APPLICANT: Patturajan, Meera
APPLICANT: Casman, Stacie J
APPLICANT: Boldog, Ferenc
APPLICANT: Gusev, Vladimir Y
APPLICANT: Burgees, Catherine E
APPLICANT: Edinger, Shlomit R
APPLICANT: Gangolli, Esha A
APPLICANT: Malvankar, Uriel M
APPLICANT: Gunther, Erik
APPLICANT: Smithson, Glenda
APPLICANT: Millet, Isabelle
APPLICANT: Gerlach, Valerie
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
TITLE OF INVENTION: Using the Same
FILE REFERENCE: 21402-221
CURRENT APPLICATION NUMBER: US/10/023,634
CURRENT FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: 60/256,025
PRIOR FILING DATE: 2000-12-15
PRIOR APPLICATION NUMBER: 60/265,163
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/272,929
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/274,864
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/276,688
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/277,880
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 60/286,409
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/309,246
PRIOR FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/315,600
PRIOR FILING DATE: 2001-08-29
NUMBER OF SEQ ID NOS: 132
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 93
LENGTH: 5636
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (3003)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
OTHER INFORMATION: specification.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (3041)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
OTHER INFORMATION: specification.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (3367)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
OTHER INFORMATION: specification.
US-10-023-634-93

Query Match 71.5%; Score 2614.8; DB 14; Length 5636;
Best Local Similarity 9.5%; Pred. No. 1.5e-29;
Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S-----I-----SS---D-----EV-----8
DB 1 MISWEVHTVFLPALLYSLAQDASPQSEIRAEPEFGASTLAFVFDVTGSMYDDLQVI 60
QY 9 -----NF-LV-----12
DB 61 EGASKILETSLRPKPLNFALPHDPEIGPVTTITDPKQFQELRELYVGGGDCPE 120
QY 13 -----Y-----R---Y-----LQ-----17

DB 121 MSIGAIIKIALEISLPGSFYIVFTDARSKOYRLTHEVLQIIQQKQSQVFLVLTGDCDDRTH 180
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
DB 181 IGKVVYEEIATSSGGVFHLDKKQVNEVLKQVVEAVQASKVHLLSTDHLEQAVNTWRIPF 240
QY 27 -----T-----FG-----I-----K 31
DB 241 DPSLKEVTVLSGSPSPMIEIRNPLGKLIKKGFGHELLNIHNSAKVNVNKEPEAGMTVK 300
QY 32 ---S--H-----I-----S--Q-----36
DB 301 TSSGRHSVRITGLSTIDFRAGSRKPTLDFKTKTSRVPVQGIPTYYLLNTSGISTPARID 360
QY 37 -----S-----NI-----N-----G-----41
DB 361 LLELLSISGSSSLKTIPIVKYIPHRKPYGIWNIISDFVPPNEAFFLKVTGYDKDYLFORVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
DB 421 VSFSSIVPDAPKVTMEKTPGVYLPQGPICSVDSLLPFTLSFVRNGVTGLGVQYLKESA 480
QY 48 -----A--L-----I-----S-----IIQ-----KG-----56
DB 481 SVSLDIKVTLSDEGFYECIAVSSAGTGRAOTFFDVSEPPVVIQVPPNVTVTGERAVLT 540
QY 57 -----L--Q-----Y--V--E-----61
DB 541 CLIIISAVDYNLTWQRNDROVRLAEPARTLANLSLELKSVEKFNDAAGEYHCVMSSEGSS 600
QY 62 -A-----EVS-----NE-----68
DB 601 AASVFLTVQEPKVTWPKNQSFSGSEVSIKMSATGYPKPKIATWVNDMFIVGSHRYM 660
QY 69 ---DGLTF-----D-----G-----R--P-----I--78
DB 661 TSDGTLFIKNAAPKAGIYGCLASNSAGTDKONSTLRYIEAPKLMVQSELLVALGDITV 720
QY 79 -E--S-----L--S--LI-----DA--V-----87
DB 721 MECKTSGIPPPQVKWPKGDLERLPSTFLIIDPLGLLKIQETQDLQDAGDYTCVAINAAGR 780
QY 88 -----M-----P--DV-----VQ-----T-----94
DB 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMFIISRPF 840
QY 95 -----R-----95
DB 841 VSSISQRLTGALFILNLWASDKGTVICEAENOFKIQSETTVTVTGLVAPLIGISPSVAN 900
QY 96 ---Q-----A-----Y-----R--D--KL--103
DB 901 VIEGQQLTLPCTLLAGNPPIERRWIKNSAMLLQNPIYTVRSOGSLHIERVOLQDGGEYTC 960
QY 104 ---A-----Q--Q-----AA-----109
DB 961 VASNVAGTNKTTSVVHVHLPITQHQQILSTIEGIPVTLPCCKASGNPKPSVIWSKKGEL 1020
QY 110 ---AAA--A--A-----A--AA--A-----118
DB 1021 ISTSAKFSAGADGSLVYVSPGESESGEYVCTATNTAGAKRKVQLTVYVPRVFGDLRG 1080
QY 119 -SQ-----Q-----GS-----123
DB 1081 LSQDKPVEISVLAGEBVTLPCEVKSLPPPIITWAKETQLISPFSPRHTFLPSGSMKITET 1140
QY 124 -----A-----K-----N-----126
DB 1141 RTSDSGMVLCVATNAGNVTAQKLVNHHVPPKIQRPKHLKVQGVQGVQVPCNAOQTPLP 1200
QY 127 -----GB--N--T-----A-----N--G-----133
DB 1201 VITWSKGSMTMLVDGSHHVNPDGTLSDQATPSDAGIVTCVATNIAGTDEBTILHVOE 1260

QY 134 -----E-----E-----N-----GA-----H----- 139
Db 1261 PPTVEDLEPPYNTTFOERAVANQRIEFPCPAKGTPTKPTIKWLHNGRELTGTEPGISILEEG 1320
QY 140 T---IA-----N-----N-----HT-----DM----- 148
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKDKQEVSNVSVLLNQLTNL 1380
QY 149 -----M----- 150
Db 1381 FCEVEGTPSPIMYKONVQVTESTSTIQTNNNGKILKILFRATPDACKYSCKAINIAGTS 1440
QY 151 -----VD-----G-----DV-----E-----I-----P-----PN 160
Db 1441 QYFNIDVLVPTTIIGTNFPKESVVLNRDVALSCQVKGTFFPDIIHWFKDGKPLFLGDPN 1500
QY 161 -----K-----A-----V-----VL 165
Db 1501 VELLDRGOVLHLKNARNDKGRYQCTVSNAGKQAKDIKUTIYIPPSIKGGNVTTDISVL 1560
QY 166 -----RG----- 168
Db 1561 INSLIKLECKTRGLPMPALITWYKDGQPTMSSQALYIDKGQYLHPRAQVSDSATYTCV 1620
QY 169 -----E-----S-----E-----VFI-----C-----A-----W----- 177
Db 1621 ANVAGTAKSPHVDVYVPPMTEGNLATPLNKQVVIASHLTLTLECNAGNPPSILTLWKDGV 1680
QY 178 -----N----- 182
Db 1681 PVKANDNFRIEAGCKLEIMSAQEI DRGOYICVATSVAGEKEIKYEVVDVLPVPAIEGDE 1740
QY 183 -----LL-----ASGS-----G-----D-----S----- 191
Db 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKDGQLIDERDGFKILLNGKLIVIAQAQVSN 1800
QY 192 -----TA----- 196
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSGSLSERVVVVKYKVALQCIANGIPNPSITW 1860
QY 197 -----NL----- 199
Db 1861 LKDPQPVNTAQNLKIQSSGRVLQIAKTLLLEDAGRYTCVATNAAGETQQHQLHVHPPS 1920
QY 200 E-----N-----S-----T-----SGST----- 207
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSGSTSTMTFLNRGQIIDIES 1980
QY 208 Q-----L-----V-----LR-----H-----C-----I----- 215
Db 1981 AQISDAGIYKCVAINSAGATELFYSQVHVAPSISGNNMVAVVNNPVRLCEARGIPA 2040
QY 216 -----R-----E-----G-----G-----Q-----D----- 221
Db 2041 PSLTWLKDGPVSVSSFNGLQVLSGRILALTSTQISDTGRYTCVAVNAAGEKQDIDLVR 2100
QY 222 VPSN----- 231
Db 2101 YVPPNIMGEEQNVSLISQAVELLCSDAIPPTTLTWLKDGHPLKPKGLSISENRSVLK 2160
QY 232 -----D----- 240
Db 2161 IEDAOVDGTGRYTCVATNAGTKEKNYNNYNIWPPNIGGSDTLTQLTVEGNLISLCE 2220
QY 241 AT-----GS-----Y-----D-----G-----F-----A----- 249
Db 2221 SSGIPPNLIWKKGSPVLTDMSGRVILSGGRQLQISIAEKSDAALYSVAVNAGTAK 2280
QY 250 -----R----- 256
Db 2281 KEYNLQYIRPTITNSGHPTEIIVTRKGSISLECEVOGIPPTVTVMKDGHPLIKAGV 2340

QY 257 -----NL-----AS-----T-----L-----GOH----- 265
Db 2341 EILDEGHILQKNHSDTGRYVCVAVNVAGMTKKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 K-----G-----PI-----FALK-----W----- 274
Db 2401 EKNSVSLTCEASGIPLPSTIWF--KDGWPVLSLSNVRILSGGRMLRLMQTTMEDAGQYTC 2458
QY 275 -----N-----K-----K-----GN----- 279
Db 2459 VVERNAAGEERKIFGLSVLPVPPHIVGENTLEDVKVEKQSVTLTCEVTGNPVEITWHKDG 2518
QY 280 -----F-----I-----S-----AGV----- 286
Db 2519 QPLODEAHHIISSGRFLQITNVQPHGTGRYTCCLASSPAGHKSRSFSLNVFVSPTIAGVG 2578
QY 287 D-----K-----T-----I-----I-----W----- 293
Db 2579 SDGNPEDVTILNSPTSLVCEAVSYPPATITWFKDGTPLSNRNIRILPGRPTLQILNAQ 2638
QY 294 D-----A-----H-----TG-----E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIIINKDGLWGPGLSPKEVKIKVNNTLLECE 2698
QY 300 A-----K-----Q-----Q----- 303
Db 2699 AYAI PASASLSWYKDGOPKSDDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758
QY 304 F-----P-----F-----H-----H 307
Db 2759 FQVNIQVPSFQKWEIGNMLDTRNGEAKDVIINNPI SILYCEETNAAPPTLTWYKDGHP 2818
QY 308 SA-----P-----A-----LDV-----D 315
Db 2819 LTSSDKVILIPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878
QY 316 -----WO-----SN----- 319
Db 2879 LPEVTVLVNKSALIECLSSGSPAPRNSWQKQDOPLEDHDKHFLNGRILQILNTQITD 2938
QY 320 -----NT-----FAS-----C-----S 326
Db 2939 IGRYVCVAENTAGSAKKYFNLNVHVPVPSVIGPKSENLTVVNNFISLTCEVSGFPDPDL 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKNXQPIKLNNTLIVPGGRTLQIIRAKVSDGGEYTCIANKAGESKKFSLTVVPPS 3058
QY 332 H-----V-----C-----K-----L-----GO----- 338
Db 3059 IKDHDSESLSVNVREGTSVLECESNAVPPPVITWYKGRMITESTHVEILADQOMLHI 3118
QY 339 D-----R-----P-----IK----- 343
Db 3119 KKAESVDTQYVCRAINVAGRDDKNFNLNVVPPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 T-----P-----Q-----G-----H-----T----- 349
Db 3179 IPPPTTIAWLKNYKRIENSLSLEVRILSGGSKLQIARSQHSDSGNYTCIASNMEGKAQKY 3238
QY 350 -----N-----E-----V-----NA-----IKW-----D-----P-----T----- 360
Db 3239 FLISQVPPSVAGNEIPSDSVLLGENVELVCHANGIPTPLIQWLKDKGFIASGETERIRV 3298
QY 361 G-----NLL-----A-----S-----C-----S-----D-----D-----MTL----- 373
Db 3299 SANGSTLNIYALGALTSDTGRYTCVATNPAGEEDRIFNLVVYVTPTRGNKDEAKLMTYVD 3358
QY 374 K-----K-----I-----W-----SM-----KQ-----D-----NCV----- 384
Db 3359 TSINIECRXTGTPPPPOINLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVSN 3418
QY 385 -----H-----D-----L-----Q----- 388

Db 3419 RAGVDNKHYNLOVAFAPNMDSNGTBEITVLKSGSTSMACITDGTAPASMAWLRDQPLG 3478
QY 389 - - - - - Q - - - - - H-N - - - - - KE 393
Db 3479 LDAHLYTSTHGMVLQLLKARTDSGKYCTIASNEAGEVSKHFLKVLPPHNGSEHEE 3538
QY 394 I - - - - - Y - - - - - I - - - - - K - - - - - W - - - - - 399
Db 3539 ISVIVNNPLELTCIASGIPAPKWTMKDGRPLQTDQVQLGGGEVLRISTAQVEDTGRY 3598
QY 400 - - - - - SPTG - - - - - P - - - - - GT - - - - - N - - - - - 407
Db 3599 TCLASSPAGDDKEYLVRHVHPNIAGTDEPRDITVLENRQVTLECKSDAVPPPVITWLR 3658
QY 408 N - - - - - P - - - - - N - - - - - ANLM-LAS - - - - - A - - - - - S - - - - - F - - - - - 420
Db 3659 NGERLOATPRVILSGRYLOINNADLGDGTANYTCVASNIAGTKTREFILTVNVPNIK 3718
QY 421 - - - - - DSTV - - - - - RL-W - - - - - D - - - - - V - - - - - DR - - - - - G - - - - - I - - - - - 433
Db 3719 GPQSLVILLKSVLECIASGVPTPRITWRKDGAVLAGNHARYSILENGFLHIOAHVTD 3778
QY 434 - - - - - C - - - - - I - - - - - H - - - - - I - - - - - TL - - - - - T - - - - - 439
Db 3779 TGRYLCMATNAAGTDRRRIDLVHVPPSIAPGPTNMTVIVNQVTLTACEATGIPKPSINW 3838
QY 440 K - - - - - H - - - - - Q - - - - - E - - - - - P - - - - - V - - - - - Y - - - - - 446
Db 3839 RKNGLHLLNDQNSRYLLSSGLVLIISPSVDDTATYECTVTNGAGDDKRTVLDLTQVPP 3898
QY 447 SVA - - - - - I - - - - - F - - - - - S - - - - - P - - - - - D - - - - - 453
Db 3899 SIADPTDLVTKHAPAVICTASGVFPFPIHWTGKIRLLPRGDGYRILSSGAIBLAT 3958
QY 454 - - - - - GRV - - - - - I - - - - - L - - - - - ASG - - - - - S - - - - - 461
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIOPOPSSELHVLNPNILLPCEATGTPSP 4018
QY 462 F - - - - - D - - - - - D - - - - - KCV - - - - - 466
Db 4019 FITWQEGINVTSGRNHVLPSGGLQISRAVEDAGTYMCVAQNPAGTALGKILNVQV 4078
QY 467 - - - - - H - - - - - I - - - - - W - - - - - 469
Db 4079 PPVISHLKEYVIADVDPITLSCADGLPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 470 - - - - - N - - - - - T - - - - - Q - - - - - 472
Db 4139 VQPDAGHYTCMAANVAGSSSTKLTVHVPPIRSTEGHYTVNENSQAILPCVADGPT 4198
QY 473 - - - - - T - - - - - G - - - - - A - - - - - L-VH 478
Db 4199 PAINWKNVLLANLLGKYTAEPYGBELILENVVLEDSGFYTCVANNAGEDTHVSLTVH 4258
QY 479 - - - - - S - - - - - Y - - - - - R - - - - - 481
Db 4259 VLPTFTELPDGSVNLKGEQLRSLKATGIPLPKLTWTFNNIIPAHFSDVNGHSELVIER 4318
QY 482 - - - - - GT - - - - - G - - - - - IF - - - - - E - - - - - V - - - - - C - - - - - 490
Db 4319 VSKEDSGTYVCTAENS VGFVKATGFFVVKPPVFKGDYPSNWIEPLGGNAILNCEVKGDP 4378
QY 491 - - - - - WN - - - - - A - - - - - AGD - - - - - 496
Db 4379 TPTIOWNRKGVDIIEISHRIQLONGSLAIYGTVNEDAGDYTCVATNEAGVVERSMSTLQ 4438
QY 497 - - - - - K-V - - - - - G - - - - - A - - - - - SAS - - - - - 503
Db 4439 SPPIITLEPVETINAGGKIILNCQATGEPOPTITWSRQHSLSWDRVNVLSNNSLYTA 4498
QY 504 - - - - - D - - - - - GSV - - - - - C-V - - - - - 509

Db 4499 DAQKEDTSBEFCVARNLMGSLVLRVPVIVQVHGGFSQWSAWRACSVTCGKIQRSLCN 4558
QY 510 - - - - - L - - - - - DL - - - - - R - - - - - K - - - - - 514
Db 4559 QPLPANGKPKCQSGDLEMCNCQNK 4582
RESULT 8
US-10-408-765A-1895
; Sequence 1895, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1895
; LENGTH: 5636
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 3003, 3041, 3367
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-408-765A-1895

Query Match 71.5%; Score 2614.8; DB 16; Length 5636;
Best Local Similarity 9.5%; Pred. No. 1.5e-29;
Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S - - - - - I - - - - - SS - - - - - D - - - - - EV - - - - - 8
Db 1 MISWEVHTVFLFALLYSLAODASPSQSEIRABEPPEGASTLAFVFDVTGSMYDDLQVI 60
QY 9 - - - - - NF-LV - - - - - 12
Db 61 EGASKILETSKRKRPLNFALVPPHDPFEGVTTTDPKPFQYELRELYVGGGDCPE 120
QY 13 - - - - - Y - - - - - R - - - - - Y - - - - - LQ - - - - - 17
Db 121 MSIGAIIKIALEISLPGSFYVFTDARS KDYRLTHEVLQLIQKQSQVVFVLTGDCDDRTH 180
QY 18 - - - - - E - - - - - SG - - - - - F - - - - - S - - - - - H - - - - - S - - - - - A - - - - - F 26
Db 181 IGKVVYEETASTSGQVPHLDKKQVNEVLKWEAEVQASKVHLLSTDHLEQAVNTWRIF 240
QY 27 - - - - - T - - - - - T - - - - - FG - - - - - I - - - - - K 31
Db 241 DPSSLKEVTVSLGSPSMIEIRNPLGKLIKGFGLHLLHNSAKVNVNKEPEAGMWTVK 300
QY 32 - - - - - S - - - - - H - - - - - I - - - - - S - - - - - Q - - - - - 36
Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFFKTVSRPVQGIPTVYLLNTSGISTPARID 360
QY 37 - - - - - S - - - - - S - - - - - NI - - - - - N - - - - - G - - - - - 41
Db 361 LLELLISGSLKTIPIVKYPPHKKPYGIWNISDFVPPNEAFFLKVTGYDKDYLFORVSS 420
QY 42 - - - - - A - - - - - L - - - - - VP - - - - - P - - - - - A 47
Db 421 VSFSSIVDPAPKVTWPEKTPGYLQPGQIPCSVDSLLPFTLSFVRNGVTLGVDQYLKESA 480
QY 48 - - - - - A - - - - - L - - - - - I - - - - - S - - - - - IQ - - - - - KG - - - - - 56

300 A-----K-----Q-----Q-----303
2699 AYAI PASLSWYKGGPLKSDHVNIAANGHTLQI KEAQISDTGRYTCVASNIAGEDEL D 2758
304 F-----P-F-----H-307
2759 FDNIOVPPSFOKLWEIGTNGEAKDVII NNPI SLYCETNAAPPPTLWYKDGHP 2818
308 --SA-----P-----A-----LDV-----D 315
2819 LTSDDKVLILPGGRVLIQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPII EGANS D 2878
316 -----WQ-----SN-----319
2879 LPEVTVLVNKSALIECLSSGSPAPRNSWKDQGPLLEDDHKKFLSNRILQILNTQIT D 2938
320 -----NT-----PAS--C-----S 326
2939 IGRYVCVAENTAGSAKKYFNLNVHVPVPSVIGPKSENLT VVVNNFISLTCEVSGFPDLS 2998
327 -----T-----D-----MCI-----331
2999 WLKXQPIKLTNTLIVPGRTLIQIRAKVSDGEYTCI AINXAGESKKFSLTVVPPS 3058
332 --H-----V--C-----K-----L--GO-----338
3059 IKDHSLSLVNVREGTSVLECESNAVPPPVITWYKNGRMITESTHVEILADGQMLHI 3118
339 -----D-----R-----P-IK-----343
3119 KKAESDGTQYVCRAINVAGRDKNFHLNVVPPSIEGP EREVIVETISNPVLTCDATG 3178
344 -----T-----F-----Q-----G-----H-----T-----349
3179 IPPPTIAWLKNYKRIENSLSLEVRILSGSKLQIARSOHSDSGNYTCIASNMEGKAQKY 3238
350 -----N-E-V-NA-----IKW--D--P--T-----360
3239 FLSIQVPPSVAGAEIPSDSVLGENVELVCNANGIPTPLIQWLKDKGPASGETERIRV 3298
361 --G--NLL-A--S-----C--S-----D-----MTL--373
3299 SANGSTLNIYALTSDTGKYTCVATNPAGEEDRI FNLNVVYPTIRGNKDEAEKLT YVD 3358
374 -----K-----I-W-----SM-----KQ--D--NCV--384
3359 TSINIECRXTGTPPPQINWLKNGLP LPLSSHIRLAAGQVIRIVRAQVSDVAVYTCVASN 3418
385 -----H-----D-----L-----Q-----388
3419 RAGVDNKHYNLQVAFPNDMSMGTEBITVLKGSSTSMACITDGTAPSMANWRDQPLG 3478
389 -----Q-----H-N-----KE 393
3479 LDAHLTVSTHGMVLQLLKAETSDSGKYTCIASNEAGEVSKHFKLVLEPPHINGSBEHR 3538
394 I-----YT-----I-K-W-----399
3539 ISVIVNNPLELTCIASGIPAPKMTWMDKGRPLPQTDQVOTLGGGEVLRISTAQVEDTRY 3598
400 -----SPTG-----P--GT-----N-----407
3599 TCLASSPAGDDKXELVVRVHVPVNIAGTDEPRDITVLRNQVLTCKSDAVPPPVITWLR 3658
408 N-----P-----N-----ANLM-LAS--A--S--F-----420
3659 NGERLOATPRVILSGRYLIQINNADLGDTANYTCVASNIAGTKTREFILTVNVPNIKG 3718
421 -----DSTV-----RL-W--D--V-----DR-----G--I-----433
3719 GPQSVILLNKSTVLECIABEGVPTPRITWRKDGAVLAGNHARYSILENGFLHQSAHVTD 3778

434 --C-----I--H-----TL--T-----439
3779 TGRYLCWATNAAGTDRRRIDLQVHVPPSIAPGPTNTMTVI VNVQTTLACEATGPKPSINW 3838
440 -K--H--O-E-----P-V--Y-----446
3839 RKNGHLLANDQNSYRLSSGSLVSIISPSVDDTATYECTVTNGAGDKRTVBLTVQVPP 3898
447 SVA-----F-S-----P-D-----453
3899 SIADEPTFLVTKHAPAVITCTASGVPPFSIHWTKNIGIRLLPRGDSYRILSSGAIEIAT 3958
454 -----GRY-----L--ASG--S--461
3959 QLNHAGRYTCVARNAAGSAHRHVTLVHBPVVIQOPSELHVILNPNPILLPCEATGTPSP 4018
462 F-----D-----KCV-----466
4019 FITWQKEGINVNTSGRNHVLPSGGLQISRAVREDAGTYMCVAQNPAGTALGKILNVQV 4078
467 --H-----I-W-----469
4079 PPVISPILKXYIADVDPITLSCADGLPPPDI THWKGRAIVESIRQVRVSSGSLQIAF 4138
470 -----N-----T-----Q-----472
4139 VQPGDAGHYTCMAANVAGSSSTSTKLT VHVPPRIIRSTEGHYTVNENSOAILPCVADGIPT 4198
473 -----T-----G-----A-----L-VH 478
4199 PAINWKONVLLANLIGKYTAEPYBELILENVVLEDSGYFTCVANNAAGDTHTVSLTVH 4258
479 -----S-----Y-----R 481
4259 VLPTTFELGDSVLNKGEOQLRUSCKATGIP LKLTWTFNNIIPAHFDSVNGHSELVIER 4318
482 --GT-----G--G-----IF-----E--V--C-----490
4319 VSKEDSGTYVCTAENS VGFKATGTFVYVKEPPVFKGDPYPSNWIEPLGGNAILCEVKGDP 4378
491 --WN-----A-----AGD-----496
4379 TPTIQNRKGVDIETISHRIQLGNGSLAIYGT VNEADAGTYCVATNEAGVVERMSLTLQ 4438
497 -----K-V--G-----A-----SAS-----503
4439 SPPIITLEPVETVINAGGKIILNCOATGEPOPTITW SROGHSISWDDRVNVLNNSLYIA 4498
504 --D-----D-----GSV-----C-V-----509
4499 DAQKEDTSBECVARNLMGSLVLRVPVIVQVHGFSQMSAWRACSVTCGKGIOKRSRLCN 4558
510 --L-----DL--R-----K 514
4559 QPLPANGKPKCGSDLEMCNQK 4582

RESULT 9

US-10-464-368-69
; Sequence 69, Application US/10464368
; Publication No. US20040023356A1
; GENERAL INFORMATION:
; APPLICANT: Krumlauf, Robb
; APPLICANT: Ellies, Debra
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
; FILE REFERENCE: 40716-IP-017
; CURRENT APPLICATION NUMBER: US/10/464,368
; CURRENT FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/388,970
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 69

```

; LENGTH: 4599
; TYPE: PRT
; ORGANISM: MOUSE
US-10-464-368-69

Query Match      71.3%; Score 2610.5; DB 15; Length 4599;
Best Local Similarity 9.5%; Pred. No. 6.3e-30;
Matches 434; Conservative 56; Mismatches 16; Indels 4085; Gaps 365;

Qy 1 MS-----I-----S-----4
Db 1 MSQALLAILTSLGLPNAEVLIVGANODQLCDPGEFLCHDHVTCVQSMLCDGDPDPCPD 60
Qy 5 SD-----EV-----N-----F-----10
Db 61 QSDSLDTCPEVEIKPLNHIACHGSSACVHLSKLCNGVVVDCPDGDFEGHCOELLPS 120
Qy 11 -L-----V-----Y-----R-----Y-L--Q-----17
Db 121 QQLNCQFKCAWRNATRCYCEDGEVAEDGRSKDQDECSYGICSTCKNTYGSYACSC 180
Qy 18 -ESGF-----S-----H-----SAF-T--F--GIK-----S-----H---33
Db 181 VE-GYIMQSDNRSCVKHEPTDKAPMLLISSLETIELFYINGSKWTTLSSANRNEIHTLD 239
Qy 34 -----I-S-QS-----NI--39
Db 240 FIYSEMICWIESRESSNQLKCGOITKAGRLTDQRIINSQSOFQNEQMAFDMLTRNIYF 299
Qy 40 -----NG-----A--L-----VP-----45
Db 300 VDHVSDRIFVCFNFGSCVTLIESELENPKAIAADPIAGKLFFTDYGNVPKIERCDDGM 359
Qy 46 -----P-A-AL-----I-----SIIQKGLQ---Y--59
Db 360 NRTRIVYSKAEQPSALALDLNRLVYVWDLVDYGVGVYQGNRHTIVQ--GRQVRHLYG 418
Qy 60 -V-E--AEVS-----IN-----E-68
Db 419 ITVFEDLYIA--TSSDNFIIRINFPNGTDIHSIIKESARGIRTYQKRTQTVRSACEV 477
Qy 69 D-----G-----TL-----F-----DGR-----76
Db 478 DAYGMPGCSHICLLSSSYKTRTCRCRTGFNMGSDGRSKRPKNELFLFYGKGRPGIVRG 537
Qy 77 -----PIE-----SL-----S-LI-----84
Db 538 MDLNTKIADECMPIENLVNPRALDPHAEANYIYFADTTSLFGRQKIDGTERETILKDD 597
Qy 85 D-----AV-----M-----P-----89
Db 598 LDNVEGIADVWIGNNLTWTDNGHRTINVARLEKASOSRKTILEGGMHPRAIVDPVNG 657
Qy 90 -----D-V-----V--Q-----93
Db 658 WMYWTWKEDKIDDSVGRIEKAMWGVNRQVFTVSKMLWPNGLTLDFTSTLYWCDAAYD 717
Qy 94 -----T-R-----Q-----96
Db 718 HIEKVLNGTHRKVVYSGKELNHPFGLSHGNVYVFTDYMGNSIFQLDLMTNEVTLRHE 777
Qy 97 -----QAY--R--D-K-----LA-----104
Db 778 RAPLFGLOIYDPRKQGDNMCRINNGCGGLTCLIAIPAGRVACADNQLLDENGTTCTFNP 837
Qy 105 -----Q-----Q-----Q-----107
Db 838 EEIFHICKPGEFRCKNKHCIQARWKCDGDDCLDGDSEDSVTCFNHSCPDQFKQNNR 897
Qy 108 -----A-----AA--AA-----AA-----AA--116
Db 898 CIPKRWLCDGANDCGSNEDESNOTCTARTCOADQFSCGNRCIPTAWLCLDREDDCGDQTD 957
```

```

117 -AAS-----Q-----Q-----121
958 EVASCEPPTCEPLTQFICKSGRCISNKHCHCDTDDCGDRSDEVCVHSCLDQFRCSSGR 1017
122 ---GS-A--KN--G---E---N-T---A-----N-----GE-134
1018 CIPGWACDGDNDGDFSDETHINCTKEARSAPAGCIGNEFQCRPGDNCIPDLWRCDGEK 1077
135 ---E---NGA---H---T---I-----141
1078 DCEDGDEKGCNGTIRLCDHKTKFSCRSTGRCINNAWVCDGDVDCEDQSDDEDCDCLG 1137
142 ---AN-----NH-----145
1138 PPKYPCANDTSCVLOPEKLCNCRKDCPDGSDGDEGLDCECLNNGGCSNHCSSVPGRGIVC 1197
146 ---T---D-----MM-----E-----VD-152
1198 SCPEGHQKKDNRTCEIVDYCASHLRCSQVCEQKHMKVCSYEGWALGTDGESCTSDS 1257
153 -----GD-----154
1258 FEAFIIFSIHRIRRIDLHKGDYSLVPLRLNTIALDPHFNOQLLYWTDVVEDRIYRGKL 1317
155 -----VE-----I-----157
1318 SESGGVSAIEVVVEHGLATPEGLTVDWIAGNIYWDNSLDQIEVSKLDGSLRATLAGAM 1377
158 --P---P-----N-----K-----161
1378 EHPRAIADPRYGILFTWDANDANFPRIESASMSGAGRKTIYDKMKTGWPNGLTVDHFER 1437
162 -----AV-----VLRGE-----SEV-----172
1438 RIVMTDARSDAIYSAFYDGTNMTIEIRGHEYLHPFAVSLYSEVYTWDMRTNTLAKANK 1497
173 -----F-----I-----CA--176
1498 WTCQNVSVIQTSAQPFDLQIYHPSROPQPNPCAANEGPGCSHLCLINHNRSACACP 1557
177 -----W-----NP-----VS--D---182
1558 HLMKLSDDKTCYEMKFLLYARRSIRBGVIDNPVNFITAFITPDIDDAVIDPDASE 1617
183 --LL-----A--SG-----S--G--D-----S--T---192
1618 ERLYWTDIKTITRAFINGTGLTETVISRDIOSIRGLAVDWVSRNLYWISSEFETOINV 1677
193 AR-----I-----W-----N-----197
1678 ARLDGSLKTSIIHGIDKPOCLAAHPVRGKLYWTDGNTINMANMDGNSKILFQKQKEPVG 1737
198 LS---EN---S--T---SG-----S-----T-----207
1738 LSIIDYVENKLYWISSNGTINRCNDGNNLEVIEMKEELTRATALTIMDKLWADQNL 1797
208 -QI-----VLR-----H-----213
1798 AOLGTCKNRDGRNPSILNRKTSVGHVHMVYDKAQQGSNSCQVNNGGCSQLCLPTSETTR 1857
214 -----C-----IR-----216
1858 TCMCTVGYLYQKRMSCOGIESFLMYSVHEGIRGLEPRDKVDALMPISGAFAVADIF 1917
217 -----EG-----218
1918 HAENDTIYWTMGLNKISRAKRDQTKEDVVTNGLGRVEGIAVDWIAGNIYWTDHGFLNI 1977
219 -----GO-----D-----V 222
1978 EVARLNGSPRYVIIISGLDQPRSIAPHPEKGLFTWEGQVPCIGKARLDGSEKVMIVSV 2037
223 --P--S-----NK-----D---V---T-----S-LD-----232
```

Db	2038	GITWPNGISIDYEENKLYWCDAKSDKIERIDLTGARNREVLISGSNDLFSVAVFCAYII	2097	Db	3114	SKLNGLYPTVLVSKRLKFPRLDLSLPRAGNLYWIDCCBYPHIGRVMGMDGTNOSVVIETKI	3173
Qy	233	W-----N-S-----	238	Qy	368	S-----D-----D-----M-----	376
Db	2098	WSDRAHANGSVRRGHKNDAETETVMTGLGVNLKEIKIFNRVREKTNVCAKENGCCOOL	2157	Db	3174	SRPMALTTIDYVNRHLYWADENHIEFSNMDGSHRHKVPNQODIPGVIALTLFEDYIYTDCGK	3233
Qy	239	-----LLA-----T-----G-----	245	Qy	377	-----S-MK-----Q-D-----N-----C-----	383
Db	2158	CLYGRNSRRTCAHGYLAGDGVTCRLRHEGYLLSGRTILKSILHSDETNLNSPVRPYEN	2217	Db	3234	TKLSRVHKTSGADRLSLNSHWAITDIQVHVSYPQDVSKHLCTVNNGCCSHLCLLPG	3293
Qy	246	-----D-----G-----F-----A-----R-----I-----	251	Qy	384	VH-----D-----L-----	387
Db	2218	PNYFKNIIALAFDYNORREGTNRIFYSDAHFGNIQLIKDNWEDROVIVENGSVEGLAYH	2277	Db	3294	KTHTCACPTNFYLAADNRNCTLSNCTASOFRCKTDKICIPFWKCDTVDDCGDGSDEDDCP	3353
Qy	252	-----W-----TK-----D-----G-----	258	Qy	388	-----Q-----Q-----H-----NKE-I-----	394
Db	2278	RAWDLYTSSSTSSITRHTVDTQTPGAIDREAVITMSDDHDPHVALDECONLFWTWM	2337	Db	3354	EFKCOPGRFQCCTGLCALPAFICDGENDCGNSDELCDTHVCLAGQFCKTKNKKICIPVN	3413
Qy	259	-----A-----ST-----	264	Qy	395	-----Y-----T-----I-K-W-----SP-----	401
Db	2338	NEQHPSIMBATLTGKNAHVVSVDILTTPNGLTIDHRAEKLYFSDGSLGKIERCEYDGSOR	2397	Db	3414	LRCNGDDCGDEDEKDCPENS CSPDYFOCKTTKHICISKLWVCDDEDDPCADASDEANCDK	3473
Qy	265	H-K-GP-----IP-ALKW-----	277	Qy	402	T-GP-----G-----T-N-----N-P-----	409
Db	2398	HVIVSGPGCTFLSLAVYDSYIFWS-DWGRRAILRSNKYTGGETKILRSIDIPHQPMGIITAV	2456	Db	3474	KTCGPHEFOCKNNNCIPDHWRCDNQDCSDNDEDNCKPQTCTLDKDFLCNSGDCVSSRFW	3533
Qy	278	-----I-----W-----D-----D-----AH-----	296	Qy	410	-----N-----A-----N-----	412
Db	2457	ANDTNSCELSPCALLNGGCHDLCLLTPDGRVNCRCGRVLLANNRCVTNKSNCIYSEF	2516	Db	3534	CDGEFDCADGSDKNCETSCSKDQFCNSGQCLSAKWKCDGHEDCKYGEDEKNCCEPAPV	3593
Qy	281	-----I-----L-----S-AG-----V-----	286	Qy	413	-----LMLA-----SAS-----F-----	420
Db	2517	ECNGDCVDYVLTCDGIPHCKDSEKLLYCENRSCRSFKCYNRCVPHGKLCDGTDND	2576	Db	3594	CSSSYMCASGCLSLKNGEPCDVGSDGDMDCVIECKEDQFOCKNKAYCIPIRWLCD	3653
Qy	287	-----D-K-----T-----T-I-----	291	Qy	421	-----D-S-----T-----V-R-----L-W-----D-V-----	429
Db	2577	CGDSDDELCKVSTCTVFRCADGTCPISARCQNQNMDCSDASDEKGCNNTDCTHFYKL	2636	Db	3654	GIYDCVDSDETCGRGSGICRDBFLCNLSLCKLHFWDCGDDCGDSDSDEAPMCVKF	3713
Qy	292	-----I-----W-----D-----D-----AH-----	296	Qy	430	-----DR-----GI-----	433
Db	2637	GVKSTGPIRCNSTSLCVPSWICDGSNDGCDYSDCLKPCVQNKHCENYFCGPGRCIL	2696	Db	3714	LCPTRPYRCNRDRIQLQLEKICNGINDCGNSDEHSCGKLSKSKPKCKDEFTCSNRN	3773
Qy	297	T-----G-E-----A-----K-----	301	Qy	434	CI-----HT-----	437
Db	2697	NTWVCDGQKDCEDGLDELHCDSSCNWQFACSVKCKISKHWICDGEDDCGDSLDESISIC	2756	Db	3774	CIPMELQDLSDDCGDSDGDEQCLKTPTEHTCENNGNPGDDAYCQIKTSVFCRCKPGF	3833
Qy	302	-----Q-----Q-----F-----	304	Qy	438	-----L-T-----K-----	440
Db	2757	GAVTCAADMFCQGGSHACVPOHWCDEGDCRDCPDGSDLSAGCAPNNTCDENAFMCHNKV	2816	Db	3834	QRNMKGRECADLINECLLFGICSHHCLNTRGSYKVCQDNQFQBNKNSCTAKSDEQALYIA	3893
Qy	305	P-P-P-H-----S-----AP-----ALD-----VD-----WQ-----	318	Qy	441	-----HQ-----E-----P-V-Y-----	446
Db	2817	CIPKQFVCDHDDCGDSDGDEFLOQGYRQGPBEFPCA-DGRCLVNTLMQCDGDFPCDPSS	2875	Db	3894	NDTDILGFVYPNYSGGHQIISHVEHNSRITGMDVHYQBNVILWSTQFNPFGIFYKNIDA	3953
Qy	319	N-----NT-FASC-----STD-MC-I-----HV-C-----	334	Qy	447	-----S-----VA-----FS-----	451
Db	2876	DEAPINPRCSAEHSCNSFFM-CXNGRCIPS-DGLCDIRDDCGDSDGDETNCHINECLSK	2933	Db	3954	REKQANSGLICPEKPRPRDIAVDWAGNVYWTDSHMHWFYSYTTHTWTSRLVSYNVGQL	4013
Qy	335	KL-G-QD-R-PI-----KT-----F-----Q-----G-----	347	Qy	452	-----P-----DG-----	454
Db	2934	KISGCSQCDQLPVSYKCKWPGFQKDDGKTCVDIDECSSGFPSCQICNTYGTVKCHC	2993	Db	4014	NGPNCTRLTNNMAGEPIAYAVNPKRGMWYTVIGDHSHEIAAAMDGTLRRVLVQKNLQRP	4073
Qy	348	-----H-----H-----T-----N-----	350	Qy	455	-----R-Y-----LA-S-----GS-----	461
Db	2994	ABGYETQDPNPGCRSLSDPEPFLILADQHRIKISTDGSNTYLLKQGLNNVALDFDYR	3053	Db	4074	TGLTVDFGERIYADWDFELSIIGSVLYDGSPPVSVSSKQGLLHHRIDVDFEDYIYGAP	4133
Qy	351	E-----V-NA-----I-K-----W-D-----	358	Qy	462	-----F-----DK-----	464
Db	3054	EEFYIWDSSRRPNSRINRCNLGSDIKVYHTATPNALAVDWIGKLYWSDEKRIIEV	3113	Db	4134	KNGIFRVQKFGHGSVEVLALGVDTKTSILVSHRYKQLNLPNFCLDLSCDFLLNPSGAT	4193
Qy	359	-----PT-----GNLL-ASC-----	367	Qy	465	CV-----H-----I-----W-----N-----	470
				Db	4194	CICPEGYMMNGTCHDDSLDDSKLCTCENGRCILNEKGDILRCHWPSYSGRCEVNH	4253

QY 162 -----AV-----VLRGHE-----SEV----- 172
Db 1438 RIVMTDARSDAIYSAFYDGTNMTIEIRGHEVLSHPPFAVSLYSEVYWTDRWTLAKANK 1497
QY 173 -----F-----I-----CA-- 176
Db 1498 WTCQNVSVIOKTSQAOPFDLQIYHPSRQOPAPNPCAANEGRGPCSHLCLINHRSAACACP 1557
QY 177 -----W-----NP-----VS--D-- 182
Db 1558 HLMKLSDDKTCYEMKKFYARRSEIRGVDIIDNPYNFITAFTVPDIDDAVIDFDASE 1617
QY 183 --LL--A--SG--S--G--D--S--T-- 192
Db 1618 ERLYWTDIKTQITRAPINGTLETVISRDIQSIGRLAVDWVRNLYWISSEFDETOINV 1677
QY 193 AR-----I-----W-----N----- 197
Db 1678 ARLDGSLKTSIIRGIDKPOCLAAHPVRGKLYWTDGNTINMANMDGNSKILFQNKPEVG 1737
QY 198 LS--EN--S--T--SG--S--T-- 207
Db 1738 LSIYVENKLYWISSNGTINRNLGGNLEVIEMKEELTKATALTIMDKKLWADQNL 1797
QY 208 -QL--VLR--H-- 213
Db 1798 AOLGTCKRDRNPISILRNKTSGVHMKVYDKEAQOQSGNSCQVNGGCSQLCLPTSETTR 1857
QY 214 -----C-----IR----- 216
Db 1858 TCMCTVGYVLQKNRMSQGIYESFLMYSVHGEIRGIPLEPRDKVDALMPISGAFAVGIDF 1917
QY 217 -----EG----- 218
Db 1918 HAENDIYWTDMGLNKISRAKRDQTKBEDVVTNGLGRVEGIAVDIAGNIYWTDHGFNLI 1977
QY 219 -----GQ-----D-----V 222
Db 1978 EVARLNGSPRYIISOGLDQPSRIAVHPEKFLFWTEWQVPCIGKARLDGSEKWLVSU 2037
QY 223 --P--S--NK--D--V--T--S--LD-- 232
Db 2038 GITWPGISIDYBENKLYWCARSDKIERIDLTGANREVLVSGSNVDLFSVAVFGAYIY 2097
QY 233 W-----N--S-----E--GT----- 238
Db 2098 WSDRAHANGSVRRGHKNDAETVTMTRTGLGVNLKEIKIFNRVREKGTNVCAKENGCCOQL 2157
QY 239 -----LLA--T-----G-----S--Y-- 245
Db 2158 CLYGRNSRRTACAHGYLAGDGVTCRLRHEGYLLYSGRITLKSITHLSDETNLNSPVRPYEN 2217
QY 246 -----D--G--F--A-----R--I-- 251
Db 2218 PNYFKNIALADYNORREGTNRIFYSDAHFGNIQLIKDNWEDRQVVENGVSGEGLAYH 2277
QY 252 -----W-----TK--D--G-----NL----- 258
Db 2278 RAMDTLYWTSSTSTSTRHTVDTQTRFCAIDREAVITMSDDHPHVLALDECQNLMEFTNW 2337
QY 259 -----A--ST-----LG-----O-- 264
Db 2338 NEQHPSIMRATLTGKNAHVWVSDIITPGLTIDHRAEKLYFSDGSLGKIERCEYDGSQR 2397
QY 265 H--K--GP-----IF--ALKW-----NK-----K----- 277
Db 2398 HVIVKSGPGTFLSLAVTDSYIFWS--DWGRAILRSNKYTGTETKILRSIDIPHQMGIIV 2456
QY 278 -----G-----N-----F 280
Db 2457 ANDTNSCELSPCALLGGCHDLCLLTPDGRVNGSCRGDRVLLANNRCVTNKSNCITYSEF 2516

QY 281 -----I-----L-----S--AG-----V----- 286
Db 2517 ECGNGDCVDVYLTCDDGI|PHCKDKSDEKLLYCNRSRSGFKPCVNRRCVPHGKLCDCGTND 2576
QY 287 -----D--K-----T-----I----- 291
Db 2577 CGDSSDELCKVSTCSTVEFRCADGTCIPRSARCNQNMDCSDASDEKGCNNTDCTHFKYL 2636
QY 292 -----I-----W-----D-----AH----- 296
Db 2637 GVKSTGFIRCNSTSLCVLPSWICDGSNDGDIYDELKCPQVQKHKCEENYFGCPSGRCIL 2696
QY 297 -T-----G--E-----A-----K----- 301
Db 2697 NTWVCDGXDCEDGLDELHCDSSCWNQFACSVKCKISKHWICDGEDDCGDSLDESISIC 2756
QY 302 -----Q-----Q-----F----- 304
Db 2757 GAVTCAADMFSQCGSHACVPQHWCDEGRDCPDGSDLSAGCAPNNTCDENAFMCHNKV 2816
QY 305 --P--F--H-----S-----AP-----ALD--VD--WQ-----S 318
Db 2817 CIPKQVCDHDDCGGSDGDEFLOCGYRQCGPEFRCA--DGRCLVNTLWQCDGDFCDPSS 2875
QY 319 -----N-----NT--FASC-----STD--MC--I-----HV--C-- 334
Db 2876 DEAPINPRCSRAEHSNCSFFM--CKNGRCIPS--DGLCDIRDDCGDSDETNCHNECLSK 2933
QY 335 KL--G--QD--R--PI-----KT-----F--Q-----G----- 347
Db 2934 KISGSCQCDQLPVSUYKCKWCPGOLKDGKTCVDIDECSSGFPSCQOCINTYGTVKCHC 2993
QY 348 -----H-----T-----N----- 350
Db 2994 AEGYETQPNPNCGRSLSDDEEPLIADQHEIRKISTDGSNYTLKQGLNVALDPRYR 3053
QY 351 -E-----V--NA-----I--K--W--D----- 358
Db 3054 EEPYIWDSSRPNRSRINRMCLNGSDIKVYHNTAVFNALAVDWIGKNLYWSDTEKRIIEV 3113
QY 359 -----PT-----GNLL--ASC----- 367
Db 3114 SKINGLYPTVLVSKRLKFRDLSDPRAGNLYWIDCCEYPHIGRVGMGTNQSIVIBTKI 3173
QY 368 S-----D-----M-----TL--K--I--W----- 376
Db 3174 SRPMALTIDYVNRHLYWADENHIEFNSMDGSHRHKVPNOQDIPGVIALTLFEDYIYWDGK 3233
QY 377 -----S--MK-----Q--D-----N-----C----- 383
Db 3234 TKLSRVHKTSGADRLSLNSWHAITDIOVYHSYRQPDYVSKHLCTVNNGGCSHLCLLPGP 3293
QY 384 -VH-----D-----L----- 387
Db 3294 KTHTCACPTNFVLAADNRCTCLSNCTASQFRCKTDKCIPEFWMKCDTVDDCGDSDPDDCP 3353
QY 388 -----Q-----Q-----H-----NKE--I-- 394
Db 3354 EFKCQGRFQCGTGLCALPAFICDGENDCGNSDELNCOTHVCLAGQFKCTKNKCLIPVN 3413
QY 395 -----Y-----T-----I--K--W-----SP----- 401
Db 3414 LRCNGODDCGDEDEKDCPENSCSPDYFOCKYTKHCISKLWVWDEDPDCADASEANCDK 3473
QY 402 -T--GP-----G-----T--N-----N--P----- 409
Db 3474 KTCGPHEFOCKNNNCIPDHWRCDNQDSDNSDEDNCKPQTC|LKDPLCSNGDCVSRFW 3533
QY 410 -----N-----N-----A-----N----- 412
Db 3534 CDGEFPCADGSDENKNCETSCSKDQFQCSNGQCLSAKWKCDGHBDCKYGEDEKNCPEFPV 3593
QY 413 -----LMLA-----SAS-----F----- 420

Db 3594 CSSEYWCAGCSLASLCKNGEPDCVDGSDMDCVIECKEDQFCCKKAYCIPRWLCD 3653
Qy 421 -D-S-T-----V-R-----L-W-----D-V-- 429
Db 3654 GIYCDVGDSETCGRGSGICRDEDFLNNSLCKLHFVWCDGEDDCGNSDEAPDMCVKF 3713
Qy 430 -DR-----GI----- 433
Db 3714 LCPTRPYRCNDRIICIQLEKICNGINDGNSDEHCSGKLSLSPCKKDEFTCSNRN 3773
Qy 434 CI-----HT----- 437
Db 3774 CIPMELQDSLDDCGSDGDEGCLKTPIHTCENNGNPGCDAYCNQIKTSVFCRCRPGF 3833
Qy 438 -L-T-----K----- 440
Db 3834 QRNMKGRECADNELLFGICSHHCLNTRGSKVCQDNFQEKNNSCIAKSGEDQALYIA 3893
Qy 441 -HQ-----E-----P-V-Y----- 446
Db 3894 NDTILGFVPFNYSGGHQQLSHVEHNSRITGMDVHYQRNVVWSTQFPGGIFYKIDA 3953
Qy 447 -S-----VA-----FS----- 451
Db 3954 REKQANSGLICPEFKRPDIADVAGNVYVTDHRSRHWFSYVYTHWTSLSRYINVGQL 4013
Qy 452 -P-----DG----- 454
Db 4014 NGPNCTRLTNMAGEYAIANPRKRGMYMTVIGDHSHEBAAMDGTLRRLVQKNLQRP 4073
Qy 455 -R-Y-----LA-S-----GS----- 461
Db 4074 TGLTVHFGRIYADPELSIGSVLYDGSPPVSVSSKQGLLHPHRIIDFEDIYAGP 4133
Qy 462 -F-----DK----- 464
Db 4134 KNGIFRVQKFGHGSVEVLALGVDTKSLVSHRYKQLNPNPCLDLSCDFLCLNPSGAT 4193
Qy 465 CV-----H-----I-----W-----N-- 470
Db 4194 CICEGYMMNGTCHDDSLDDSCKLCENGGRILNEKGLRCHCWPSYSGRCVYNHC 4253
Qy 471 T---Q-----T-G-----AL-----V-----H-S----- 479
Db 4254 SNYCQNGTCIPITLGRPTICALGFTGPNCGKAVCEDSCHNGSCVVTAGNQPYCHCOA 4313
Qy 480 Y---R-----G-----T-----GG----- 485
Db 4314 DYTGRQYVCHHYCVNSBESCTIGNDGSVECVCPTRYEGPKCEIDKVCRCGGHGIINK 4373
Qy 486 -IF-----E-----VC---W----- 491
Db 4374 DNEDIFCNCTNGKIASSCOLCDGYCYNGGTCLDPETSIPVCVSTNWSGTQCPAPAKS 4433
Qy 492 -----S-V-----C-----VLDLRK 514
Db 4554 PTNYSNPVYAKLYMDGQNCNRNSLASV-DERK 4583

RESULT 11

US-10-120-801-74

; Sequence 74, Application US/10120801

; Publication No. US20030203843A1

; GENERAL INFORMATION:

; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 74
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-120-801-74

Query Match 71.0%; Score 2596.6; DB 14; Length 5175;

Best Local Similarity 9.0%; Pred. No. 2.1e-29;

Matches 454; Conservative 48; Mismatches 10; Indels 4514; Gaps 390;

Qy 1 M-----SI-----SS-----D-----E----- 7
Db 1 MGRSPWLVGLGILLALLATTCSSVNDNDKNDPTGKSLAFVFDITGSMFDDLVQVREGAAK 60
Qy 8 -V-----N-----F-L-VY----- 13
Db 61 IPKTVMAQREKLIYNYIMVFPHPDYLGEINTTDSYFMRQLSKVYVHGGDCPEKTLTG 120
Qy 14 -----R-Y-----L-OR-SG-F-----SHSAF-T 27
Db 121 ILKALQISLPSSPIYVFTDARSKDYHLEDEVLNTIQEKSSVVFVMTGCGNRTHPGFT 180
Qy 28 -----FG-I-----KS-----H-ISQ-----S-NI----- 39
Db 181 YEKIAAASFGQVPHLEKSDVSTVLEYVRHAVKOKKVLHMYEARERGTVSRNIPVDKHL 240
Qy 40 -----N-----GA-----LV-----P-P----- 46
Db 241 ELTISLGDGKDDSDNLDIVLDPDGTVDKRLYSKGGITDLKNVKILRLKDPSPGVWTV 300
Qy 47 -----A-----A-----LI-S-I----- 52
Db 301 NTSRLKHTIRVFGHGAVDKYGFAFRPLDRIELARPRVFNQDTYLLINMTGLIPPGTV 360
Qy 53 --I-----Q-----KGL-----O-Y-----VE--AE- 63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPPVPKGLFFVRVQGVDEDNEMFIAPT 420
Qy 64 -----V-----S-I-----N-----ED-----GTLF----- 73

Db	421	AIGSVIVGGPRAPWSPHOFVGRDLNLSCTBSASAYTIYWKGTGEDIIIGGFLFYHNTD	480	Db	1500	KTSWKTSDDRKRLHVPKAKITDSGVYKCVARNAAGEGSKSFQVEIVPLNLDSEYKXKV	1559
Qy	74	-----D-----G-----R--P-I-----E-----	79	Qy	179	-----PVS-----	181
Db	481	TSVWTIPELSLKDAGEYECRIVSNNGNYSVKTRVETRESPEIFGRVNSVPLGEAFLH	540	Db	1560	FAKEGSEVTILGCPVSGFPVPOINWVVDGTVVEBKKYKGATLNDGLTLHFDVSUVKQBG	1619
Qy	80	-----S-----L-----SL-I-----DA-----	86	Qy	182	-----D--L--LA-----	186
Db	541	CSTRSAGEVEIRWTRYGATVFNGPNTERNPTNGTLKIHVTRADAGVYECMARNAAGMST	600	Db	1620	NYHCVASKGNILIDIVELSVLAVPIVGBDDNLEVLGKDISLSCDLQTESDDKTTPVMS	1679
Qy	87	-----VM-----P-DV-----	92	Qy	187	-----GS-----G-----DS-----T-----A--R-----	194
Db	601	RKRLDIMEPPSVKVPQDYFNMREGVNLSCBAMGDPKDEVHWYFKGRHLLNDYKQVG	660	Db	1680	INGSERDPDNVQIPSDGHRLYITDAKPENNGKMYCRVTNSAGKAERTLTLDVLEPPVF	1739
Qy	93	Q-----T--R--O--QAYRD-----KLA-----Q-----	105	Qy	195	-----IW--N-----LS-----	199
Db	661	QDSKFLYIRDATHHDSGTTECRAMSQAQA--RUTDMLLATPPKVEIIQNMVMGRGDRV	719	Db	1740	EPVFEANQKIGNPILOCOVGTGNPKPTVIWKIDGNDVDKSWLFDESLSLLRIEKLTKG	1799
Qy	106	-----Q-----Q-----A-----A-----A-----	111	Qy	200	-----EN-----	207
Db	720	SPECKTIRGKPHKPIRWFKNKGDLIKPDDDYKINEGQHLHMGAKEDAGAYCVGENMAG	779	Db	1800	SAQISCTAENKAGTASRDFFIQNIAAPTFKNEGQDQETIFRESEITITLDCPVLGDFQITW	1859
Qy	112	-----A-----A-----A-----A-----A-----	114	Qy	208	-----Q-----L-VL--R-H--C-----	216
Db	780	KDVQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPPEIEWQKGNVLLATLNP	839	Db	1860	MKQGLPLTENDAIFTLNDTRLTLNARDHEDIYTCVANNTAGQVSKDFDVVQVLPKIK	1919
Qy	115	-----A-----A-----ASQ-----	120	Qy	217	-----EG-----G-----Q-DVP-----	223
Db	840	RYTQADGNLLITDAIEDQGQFTCIARTNYGQOSOTTLMTVGLVSPVLGHVPPEQLI	899	Db	1920	NAVVTLEINEGBEIIITCDAEGNPTPTAKWDFNQGLPKPEAVFVNNHTVVNNVTKYHT	1979
Qy	121	-----Q-----	123	Qy	224	-----SNK-----DV-----T--S--L--D-----	232
Db	900	EGQDLTLCVVLTGPKPSIWIWKDDKPVEEGTIKIEGGSLRLRGNPKDEGKYTCI	959	Db	1980	GYYKCVATNKVQAVKTIINVHVTKPRFESGLTESELTVNLTRTSITLECDVDAIGVGIS	2039
Qy	124	-----A-----	128	Qy	233	W--N-----SE--G--TL--L-----A--T-----GSY-----	245
Db	960	AVSPAGNSTLHINVLKIKPEFVYKPEGGIVFKPTISGMDEKHVAVVNSTHVDLDEGFA	1019	Db	2040	WTVNGKPLAETDGVQTLAGRFLHIVSAKTDHGSYACTVTVNEAGVATKTNFLVQVPP	2099
Qy	129	-----NT-----	133	Qy	246	-----DG-----	248
Db	1020	IPCWVGTPPPIITWYLDGRPIITPNSRDFVTADNTLIVRKADKSYGVYTCOATNSAGD	1079	Db	2100	TIVNEGVEYTVIENNSLVLPCEVTKPNPVVTWKDGRPVGDLKSVQVLSEGGQFKIVHA	2159
Qy	134	E-----	136	Qy	249	-----A-----R-----	250
Db	1080	NEQKTIIRIMNTPMISPGQSSFNWVDDLETIPCDVYGPDPKPVITWLLDDKPTGTVNE	1139	Db	2160	ETAHKGSYICMAKNDVGTAEISFDVDIITRPMIQKGIKNIVTAIKGALPFKCPIDDDKN	2219
Qy	137	G-----AH-----	144	Qy	251	-----IW-----	257
Db	1140	DGSLTIPNVNEAHRGFTCHAQNAAGNDTRTVLTVHTTPTTNAEQEKIALQNDIVLE	1199	Db	2220	FKGQIILWNLNYQPIDLEADARITRLSNDRRRLTILNVTENDEGQYSCRVKNDAGENSFDF	2279
Qy	145	-----HT--D-----MME-----V-----D--152	152	Qy	258	-----L-----A--S--TL-----QO--H-----	265
Db	1200	CPAKALPPPVLWYBEGEKIDSQILPHITREDGALVQNVKLENTGVFCQVSNLAGEDS	1259	Db	2280	KATVLPPTIIMLDKDKNKTAVEHSTVTLSCPATGKPBPDITWFKDGEAIIHENIADIIP	2339
Qy	153	-----G--DV-----EIP-----P-----NK-AV-----	163	Qy	266	-----	265
Db	1260	LSYTLTVHEPKPIISEVPGVDVVKGTFTIEIPCRATGVBEIVTWNGKIDLMKDEKFS	1319	Db	2340	NGELNGNLKITRIKEGDAGKYTCEDNSAGSVQDVNVNVTIPKIEKDGIPSDYESQQ	2399
Qy	164	V-----LR-----G-----H-----E-----S-----	170	Qy	266	-----K--G--P-----	269
Db	1320	VDNLGTLRIEADKNDIGNYCNVVTNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGR	1379	Db	2400	NERVVISCPVYARPPAKITWLKAGKPLQSKDFVKTSANGQKLYFLKRLRETSKYTCIAT	2459
Qy	171	-----E-----V--F-----	175	Qy	270	-----P-----A-----L-----K-----W-----N--275	275
Db	1380	VELKCYVEASPPASVTFWFRGIAIGDTKGYVVESDGLVIOQSASVEDATIYTCASNPA	1439	Db	2460	NEAGTKDRFKVSMVLVAPSFDEPNIVRITVNSGNPSTLHCPAKGPSPTITWLKDNAI	2519
Qy	176	-----A-----	176	Qy	276	-----K-----	276
Db	1440	GKAENLQVTVIASPDIKDPDVVTQBSIKESHFPFSLYCPVFNPLPQISWYLNKPLIID	1499	Db	2520	EPNDRYVFFDAGRQLOISKTEGSDQRYTCIATNSVGSDDLENTLEVIIPVIDGERREA	2579
Qy	177	-----W-----	178	Qy	277	-----K--G--N-----FI-----LS-----283	283
				Db	2580	VAVIEGFSSELFCDSNSTGVDEWQKDLTINQDTRLGRGDSFIQIPSSGKGMFLSARKSD	2639

QY 284 ---AG---V---DK---T---TI---I--- 292
Db 2640 SGRVTCIVRNPAEARKLFDFAVNDPPSISDELSSANIQTIVPYVVEINCVSGSPHPK 2699
QY 293 ---W---D---A---H---T---G--- 298
Db 2700 VYWLFDKPLPDAAVELTNGETLTVRSQVEHAGTYTCEAQNNGVKARKOPFLVRVTA 2759
QY 299 ---E---AK---Q--- 302
Db 2760 PPHEKEREYVARVGTMLTLCNAESSVPLSSVYVHAHDESVQGVITSKYAANEKTLN 2819
QY 303 ---Q---F--- 304
Db 2820 VTNQLDDEGFYCTAVNEAGITKFKPLVIETPYFLDQOKLYPIILGKRLTLDCSATG 2879
QY 305 ---P---F--- 306
Db 2880 TPTPTILFMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGYRTCIAENKAGRSEKDMVVEV 2939
QY 307 ---H---S--- 308
Db 2940 LLPKLSKEMINVEVQAGDPLTECPIEDTSGVHITWSRQFGDKDQGLDMRAQSSDKSL 2999
QY 309 ---A-P-A-L---D---V-D--- 315
Db 3000 YIMOATPEDADSYSCIAVNDAGBAEVQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059
QY 316 ---W---Q--- 317
Db 3060 EGIPPEISWFLDGKPILEMPGVYKQGLSLRDNKPNQEGRYTCAENKAGRABQDT 3119
QY 318 ---S---N---T---F--- 322
Db 3120 YVEISEPRVMASEVNRVVEGRQTTIRCEVFGNPEVFNWLNKDGEPYTSLLQFSTKLS 3179
QY 323 ---A-SC---S-TD---M-C-330 330
Db 3180 YLHLRETTLDGGTYTCIATNKAGESQTTDVEVLVPPRIEDBERVLOGKEGTYVWHCQ 3239
QY 331 ---I---HV---C---K 335
Db 3240 VTGRPVYVTKRNGKEIEQFNPLVHIRNATRADEGKYSCIASNEAGTAVADFLIDVFTK 3299
QY 336 ---L-GQ---D---RP---I-K-343 343
Db 3300 PTFETHETTFNIVEGSAKIECKIDGHPKPTISWLKGRPFNMDNIILSPRGDTLMILKA 3359
QY 344 ---T---F---Q--- 346
Db 3360 QRPDGLYTCVATNSYGDSEQDFKVVVYKPYIDETIDQTPKAVAGGEIILKCPVLGNPT 3419
QY 347 ---G---HT---N-E---VN---A-I-355 355
Db 3420 PTVTKRGDDAVPNDRSRTIWNVYDLKINSVTTEDAGQYSCIAVNEAGNLTHVAAEVIG 3479
QY 356 ---K---W---D---P---T-360 360
Db 3480 KPTFVRKGNLYEVIENDTTMDGVTSRPLPSISWFRGDKPVLYLDYRYSISPDGSHITI 3539
QY 361 ---G---N---L---LA---S---C-S-368 368
Db 3540 NKAKLSDGGKYICRANEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPIIS 3599
QY 369 ---D---D-M--- 371
Db 3600 GIPQPDVIWTKNGMDINWTSRVLAQNNETFGIENVQVTDQGRYTCATNRGKASHDF 3659
QY 372 ---TL---K---I---W---S---M-378 378
Db 3660 SLDLVSPPEFIHGTOPIKREGDTITLTCPILAEADIAQVMDVMSWTKDSRALDGLDITD 3719

QY 379 ---K---Q---D--- 381
Db 3720 NVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPPVIDISRNDVQPOV 3779
QY 382 ---N---C-V-H---L---Q-Q---H--- 390
Db 3780 AVNQPTIMRCVAVTGHPPPSIKWLKNGKEVTDDEINIRIVEQGVQLILRTDSDHAGKWSVCV 3839
QY 391 ---N---KE---I-Y-TI---KW--- 399
Db 3840 AENDAGVKELEMLVDVFTPEVWSKDNFIKALGETITLFCNAGSNPYPOLKWKAGGSLI 3899
QY 400 ---SP---T-G---P---GT---N--- 407
Db 3900 FDSPDGARISLKGARLDIPLHKKTDVGDYTCQALNAAGTSEASVSDVLVPPPEINRDGID 3959
QY 408 ---N---P--- 409
Db 3960 MSPRLPAQOQLTLQCLAQKFPQMRWTLNGTALTHTSTPGITVASDSTFIQINNVSLSDK 4019
QY 410 ---N-A-NLM---LA---S--- 417
Db 4020 GVYTCYAEVAGSDNLMYVNVVQAPVINSNGTKQVIEGELAVIECLVEGYPAPQVSWLR 4079
QY 418 ---A-SFDS--- 422
Db 4080 NGRVETGVQVRYVTDGRMLTIIIEARSLDSGIYLCSATNEAGSAQAQAYTLEVLVSPKII 4139
QY 423 ---T---VR-L---W---DV--- 429
Db 4140 TSPGVLTPSSGSKFSLPCAVRGYPDPPIISWTLNGNDIKDGENGTIGADGTLHIEKABE 4199
QY 430 ---D---R---G---I-C-I--- 435
Db 4200 RHLIYECTAKNDAGADTLEFPQTVIAPKISTGNRYINGSEGTETVIKEIESESEFS 4259
QY 436 ---H---T-L---T---K 440
Db 4260 WSKNGVPLPSSNNLISEDYKLIKILSTRLSDOGEYSCTAANKAGNATQKTNLNVGVPK 4319
QY 441 ---H-Q---E---P-V-Y---SVAF--- 450
Db 4320 IMERPRTQVVKHQDQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTTKKKSVIFSSI 4379
QY 451 SP---D---G--- 454
Db 4380 SPQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMNVSTNPNROTFLVSCNATGIPEP 4439
QY 455 ---R---Y---LA---SGSF---D---K 464
Db 4440 VISWMEDSNIAIONNEKYQILGTTLAINRVLPDDDG-FYHCIAKSDAGQKIATKRLIVNK 4498
QY 465 ---C---VHI---W---N--- 470
Db 4499 PSDRPAPIWVECDKEGKPKKTEYMDRGDTDDNPPOLLPKWDEVDSSNGSIAYRCMPGP 4558
QY 471 ---T---Q---T-CA---LV--- 477
Db 4559 RSRVTLLHAAPQFIVKPKNTAAIGAIVELRCSAAGPPHPTITWAKDGKLIEDSKFEIA 4618
QY 478 ---H---S--- 479
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKSSNQKNVAVITCYE 4678
QY 480 ---Y-RG-T---G---GI--- 486
Db 4679 RNQVSRGLTWYENGVPMPKNLAGIHFMMNGSLVILDTSSLKEGDLLEYTCVKVNRERRHS 4738
QY 487 ---FE---V-C---W-N---A--- 493
Db 4739 IPHLTSAFEGVEPVKTIDKVENNGSDVLDCEVTDPLTTHVVTKNDQKMLDDDAIYV 4798
QY 494 ---A---G---D---KV---G---AS--- 501

Db 4799 LPNLSVLLNVEKYDEGVKCVASNSIGKAFDDTQLNVYGGSSRRREAYKKENEDASTTTI 4858
QY 502 -----A-----S-----DG----- 505
Db 4859 TTTSPPTTTTETPLTTIIPALITLPAKQVPTDDYHEGSANDDGFPTTQDSLFEFNPPL 4918
QY 506 -----SV-----CV-----LD----- 511
Db 4919 HPEISVVNTDCAGTINENGCDVDKDKTHNLKLTGENHCPEGFAMNPHTRICEDLDECA 4978
QY 512 -----L-----R-----K 514
Db 4979 FYQPCDFECINVDGFCNCPGLGYELAEBCRDVNECESVRCBDGK 5024

RESULT 12
US-10-369-493-6859
; Sequence 6859, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6859
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6859

Query Match 71.0%; Score 2596.6; DB 14; Length 5175;
Best Local Similarity 9.0%; Pred. No. 2.1e-29;
Matches 454; Conservative 48; Mismatches 10; Indels 4514; Gaps 390;

QY 1 M-----SI-----SS-----D-----E----- 7
Db 1 MGRSPSLYGVLLGALLLATTCSVNDKNDPTGKSLAFVFDITGSMFDDLVQVREGAAK 60
QY 8 -----V-----N-----F-----L-----VY----- 13
Db 61 IPKTVMAQREKLIYIMVPHDPYLGEIINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120
QY 14 -----R-----Y-----L-----QE-----SG-----F-----SHSAP-T 27
Db 121 ILKALQISLPSPFIYVFTDARSKDYLEDEVLTNIOEKQSSVVFVMTGDCGNRTHPGFRT 180
QY 28 -----FG-----I-----KS-----H-----ISQ-----S-----NI----- 39
Db 181 YEKIAAASFQGFHLEKSDVSTVLEYVRAVKOKVHLMYAEARCGTVSRNIPVDXHL 240
QY 40 -----N-----N-----GA-----LV-----P-----P----- 46
Db 241 ELTISLGDKDDNDLVDLROPEGRVTKRLYSKEGGTIDLKVKVILRLKDPSPGVWTV 300
QY 47 -----A-----A-----A-----LI-----S-----I----- 52
Db 301 NTNSRLKHTIRVFGHGAVDKYGFASRLDRIELARPRVNLQDTYLLINMTGLIPPGTV 360
QY 53 -----I-----Q-----Q-----KGL-----Q-----Y-----VE-----AE----- 63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNWYFAGPVPPKGLFFVRVQGYDEBNYEFMRAPT 420
QY 64 -----V-----S-----I-----N-----ED-----GTLF----- 73

Db 421 AIGSVIVGGPRAFMSPHIOEFVGRDLNLSCTVESASAYTIYVVKTGEDIIGGFLFYHNTD 480
QY 74 -----D-----G-----R-----P-----I-----E----- 79
Db 481 TSVWITPELSLKDAGEYECRVISNNNGNSVSKTRVETRESPEIFGVRNVSVPLGEAAFLH 540
QY 80 -----S-----L-----SL-----I-----DA----- 86
Db 541 CSTRSAGEVEIRWTRVGAIVFNGPNTERNPTNGTGLKIHVTRADAGVYECMARNAGMST 600
QY 87 -----VM-----P-----DV----- 92
Db 601 RKMRLDIMEPPSVKVTPOVYFNMREGVNLSCAMGDGPKPEVHWYFKGRHLLNDYKYQVG 660
QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q----- 105
Db 661 QDSKFLYIRDATHHDEGTIECRAMSQAQA-RDITDMLATPPKVEIIONKMVGRGDRV 719
QY 106 -----Q-----Q-----Q-----A-----A-----A-----A----- 111
Db 720 SPECKTIRGKPHKIRWPKNGKDLIKPDDYIKINEQQLHMGAKDEADAGAYSCVGENMAG 779
QY 112 -----A-----A-----A-----A-----A-----A----- 114
Db 780 KDQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNNP 839
QY 115 -----A-----A-----A-----ASQ----- 120
Db 840 RYQLADGNLLITDAQIEDOGFTCIARTYQQSQSOTTLMTVGLVSPVLGHVPPBEQLI 899
QY 121 -----Q-----Q-----Q-----GS----- 123
Db 900 EGQDLTLSCVVLGTPKPSIVMIKODKPVVEGPTIKIEGGSSLLRLGNGPKDEGKYTCI 959
QY 124 -----A-----A-----A-----K-----N-----GE----- 128
Db 960 AVSPAGNSTLHINVLKIKPEFYKPEGGIVFKPTISGMDEKHVAVVNSTHVDLDEGFA 1019
QY 129 -----A-----A-----A-----N----- 133
Db 1020 IPCVWSGTPTTITWYLDGRPTPNSRDFTVTADNTLIVRKADKSYGVVTCOATNSAGD 1079
QY 134 -----E-----E-----E-----E-----N----- 136
Db 1080 NEQKTTIRIMTPTMISPGOSSFNWVDDLTIPCDVYGDPKPVITWLLDDKPTGVVNE 1139
QY 137 -----G-----AH-----TI-----A-----N-----N----- 144
Db 1140 DGSLLTIPNVNEAHRGFTTCHQAAGNDTRVTLTVHTTPTINAENQEKIALQNDIVLE 1199
QY 145 -----HT-----D-----D-----MME-----V-----D----- 152
Db 1200 CPAKALPPVRLWYVEGEKIDSQLPHITIREGALVQLNKLNTGVFCQVSNLAGEDS 1259
QY 153 -----G-----DV-----EIP-----P-----NK-----AV----- 163
Db 1260 LSYTLTVHEKPKIISSEVPGVVDVKGFTTIEICRATGVEVIRTWNKNGIDLKMDKKFS 1319
QY 164 V-----LR-----G-----H-----E-----S----- 170
Db 1320 VDNLGLTRIYEADKNDIGNYCNVVTNEAGTSQMTTHVDVQEPPIILPSTQNTNNAVVGDR 1379
QY 171 -----E-----V-----F-----I-----C----- 175
Db 1380 VELKCYVEASPPASVTWFRRGIAIGTDTKGYVVESDGLTVIOSASVEDATITYCKASNPA 1439
QY 176 -----A-----A-----A----- 176
Db 1440 GKAEANLQVTVIASPDIKDPDVVTOESIKESHPPFSLYCPVFNPLPQISWYLNKPLIDD 1499
QY 177 -----W-----N----- 178

Db 1500 KTSWKTSDDKRKLHVFKAKITDGSYKVCARNAAGEGSKSQVEVIVPLNLDESXYKKKV 1559
Qy 179 -----PVS----- 181
Db 1560 FAKGEBEVLGCPVSGFPVPOINWVDTGVVEPGKYKGATLSNDGLTLHFDSVSVKQEG 1619
Qy 182 -----D-L-LA-----S 186
Db 1620 NYHCAQSKGNILDIDVELSVLAVPIVGEDDNLVFLGKISLSDQLQTESDDKTTFW 1679
Qy 187 -GS-----G-----DS-----T-----A-R----- 194
Db 1680 INGSERDPNVQIPSDGHRLYITDAKPENNGKVMCRVNTSAGKAERTLTLDLVLEPPVF 1739
Qy 195 -----IW-----N-----LS----- 199
Db 1740 EPVPEANOKLIGNPIILQCOVTCNPKPTVIWKIDGNDVDKSWLFDLSLLRIEKLTK 1799
Qy 200 -----EN-----S-----T-----S-GS-----T- 207
Db 1800 SAQISCTAENKAGTASRDFFIQNLAAPTFKNEGQETIPRESEITILDCPVSIGDFQITW 1859
Qy 208 -Q-----L-VL-----R-H-----C-----IR 216
Db 1860 MKOGLPLTENDAIFLTDNTRLTILNANRDHEDIYTCVANNTAGOVSKDFDVVQVLPKIK 1919
Qy 217 -----EG-----G-----Q-DVP----- 223
Db 1920 NAVTLNINGBEILTCDAENPTTAKWDFNOGLPKFAVFNHNNHTVVVNNVTKYHT 1979
Qy 224 -----SNK-----DV-----T-S-L-----D----- 232
Db 1980 GYKCYATNKVGQAVKTIHVHVRKPRFESGLTESELTVNLTREITILECDVDDAIGVIS 2039
Qy 233 W-N-----SE-G-TL-----L-----A-T-----GSY----- 245
Db 2040 WTVNGKFLAETDGVQTLAGREFLHVSAKTDHGSYACTVTNEAGVATKTNFLVQVPP 2099
Qy 246 -----DG-----F----- 248
Db 2100 TIVNEGYYTVIENSLVLPCEVTGKPNPVVTVTKDGRPVGDLKSVQVLSGQOQKIVHA 2159
Qy 249 -----A-----R----- 250
Db 2160 EIAHKGSIYICAKNDVGTAEISFDVDIITREMIQKGIKIVITAKGALFPKCPIDDDKN 2219
Qy 251 -----IW-----T-----K-D-G-N----- 257
Db 2220 FKQIILWRNYQPIDLEAEDARITRLSNDRLTLNVTENDEGQYSCRKNDAGENSFDP 2279
Qy 258 -----L-----A-S-TL-----GQ-H----- 265
Db 2280 KATVLVPTTIIMLDKKNKTAVEHSTVTLSCPATGKPEPDIITWFKDGEAIIHENIADIIP 2339
Qy 266 -----K-G-P-----I----- 269
Db 2340 NGELNGQLKITRIKEDAGKYTCADNSAGSVEQDVNVNVTIPKIEKDGIPSDYESQ 2399
Qy 266 -----K-G-P-----I----- 269
Db 2400 NERVISCPVYARPAKITWLGKQPLQSDKFKVTSANGQKLYLFKLRETDSSKYTCIAT 2459
Qy 270 -----F-A-----L-----K-----W-----N----- 275
Db 2460 NEAGTDKRDPKVEMLVAPSFDEPNIVRITVNSGNPSTLHCPAKGSPPTITWLDGNAI 2519
Qy 276 -----K----- 276
Db 2520 EPNDRYVFFDAGRLQISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPIDGERREA 2579
Qy 277 -----K-G-----N-----FI-----LS----- 283
Db 2580 VAVIEGFSSELCDSNSTGVDVWQKDGLTINQDTRLGRGDSFIQIPSGKQMSLSARKSD 2639

Qy 284 -----AG-----V-----DK-T-----TI-----I----- 292
Db 2640 SGRYTCIVRNPAEAKRLFDFAVNDPPSISDELSSANIQTIVPPYVPEINCVVSGSPHPK 2699
Qy 293 -----W-----D-A-----H-----T-----G----- 298
Db 2700 VYWLFDKPLPDSAAVELTNNGETLKIVRSQVEHAGTYTCEAQNNVVGKARKDFLVRVTA 2759
Qy 299 -----E-----AK-----Q----- 302
Db 2760 PPHFEKERBEVVARVGDITMLLTCAESSVPLSSVYVWHAHDESQVNGVITSKYAANEKTLN 2819
Qy 303 -----Q-----P----- 304
Db 2820 VTNIQDDEGYYCTAVNEAGITTKFFKLIVITPYFLDQOKLYPIILGKRLTLDCSATG 2879
Qy 305 -----P-----F----- 306
Db 2880 TPTPTILFMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGRTYTCIAENKAGRSEKDMVVEV 2939
Qy 307 -----H----- 308
Db 2940 LLPPLSKKEWINVEVOAGDPLTLECPIEDTSGVHITWSRQFGKQGLDMRAQSSDKSL 2999
Qy 309 -----A-P-A-L-----D-----V-D----- 315
Db 3000 YIMQATPEDADSYSCIAVNDAGAEAVQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059
Qy 316 -----W----- 317
Db 3060 EGIPPEISWFLDGKPILEMPGVYTKQDLSLRIDNIPKNOEGRYTCVAENKAGRAEQDT 3119
Qy 318 -----S-----N-----N-----T-----P----- 322
Db 3120 YVEISEPPRVWASEVMRVVEGRQTTIRCEVFGNPEPVVNWLDKGBPYTSDLLQFSTKLS 3179
Qy 323 -----A-----SC-----S-----TD-----M-C- 330
Db 3180 YLHLRETTIADGGTYTCIATNKAGESQTTDDEVLPVPRIEDEVLOKQEGNTYVWHQ 3239
Qy 331 -----I-----HV-----C-----K 335
Db 3240 VTGRPVYVTVKNGKEIEIQFNPVLHIRNATRADEGKYSIASNEAGTAVADFLIDVFTK 3299
Qy 336 -----L-GQ-----D-----RP-----I-K- 343
Db 3300 PTFETHETTFFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMNDNIIILSPRGDTMLKA 3359
Qy 344 -----T-----F-----Q----- 346
Db 3360 QRFDGGLYTCVATNSYGDSEQDFKVVNTKPYIDETIDOTPKAVAGEIILKCPVLGNPT 3419
Qy 347 -----G-----HT-----N-----E-----VN-----A-I- 355
Db 3420 PTVTKRGDDAVENDSRHTIIVNNYDLKINSVTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479
Qy 356 -----K-----W-----D-----P-----T- 360
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDGSHITI 3539
Qy 361 -----G-----N-----L-----LA-----S-----C-S 368
Db 3540 NKAKLSDGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPI 3599
Qy 369 -----D-----D-M----- 371
Db 3600 GIPOPDVITKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGKASHDF 3659
Qy 372 -----TL-----K-----I-----W-----S-----M- 378
Db 3660 SLDLVSPPEFDIHGTOPTIKREGDTITLTCPIKLAEDTADQVMDVSWTKDSRALDGLD 3719

QY 379 -----K-Q-----D----- 381
Db 3720 NVDISDDGRKLTIQASLEAGLYCTIALNRAGEASLEFKVELTSPVIDISRDVQPOV 3779
QY 382 --N--C-V-H-----D--L--Q-Q--H----- 390
Db 3780 AVNQPTIMRCVATGHPFPIKMLKNGKEVTDENIRIVEGQVLQILRTSDHAGKWSV 3839
QY 391 --N--KE-----I-Y-TI-----KW----- 399
Db 3840 AENDAGVKELEWLDVFTPPVSVKSDNPIKALGETITLFCNAGSNPYQLKWAKGSLI 3899
QY 400 --SP--T-G--P-----GT-----N----- 407
Db 3900 FDSPDGAISLKGARLDIPLHKKTVDGDTYCOALNAAGTSEASVSDVLVPPPEINRDGID 3959
QY 408 -----N-----P----- 409
Db 3960 MSPRLPAQOSLTLQLAQKVPQMRWTLNGTALTHTPGITVASDSTFIQINNVSLSDK 4019
QY 410 -----N-A--NLM-----LA-----S----- 417
Db 4020 GUYTCYAEENVAGSDNLMYNDVVOQAPVINSNGTKQVIEGELAVIECLVEGYPAQVSWLR 4079
QY 418 -----A-SFDS----- 422
Db 4080 NGNRVETGVQVRYVTDGRMLTIEARSLDSGIYLCSATNEAGSAQAQVTLVLVSPKII 4139
QY 423 -----T-----VR-L--W-----DV----- 429
Db 4140 TSTPGVLTPSSGSKFSLPCAVRGPDPIISWTLNGNDIKDGENGTIGADGLHIEABE 4199
QY 430 -----D-----R-----G-----I-C-I----- 435
Db 4200 RHLIYECTAKNDAGDTLEFPVOTIVAPKISGNRYINGSEGTETVKEIESESSEFS 4259
QY 436 -----H-----T-L-----T-----K 440
Db 4260 WSKNGVPLLPNNLIFSDYKLIKILSTRLSQGEYSCTAANKAGNATQKTNLNVGAPK 4319
QY 441 -----H-Q-----E-----P-V-Y-----SVAP-- 450
Db 4320 IMERPRTQVHKGDQVTLWCEASGVQPAPITWYKONELLTNTGVDSTATTKKSVIFSSI 4379
QY 451 SP-----D-----G-----G----- 454
Db 4380 SPQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMNVSTNPROTVFLSCNATGIP 4439
QY 455 -----R-----Y-----LA-----SGSF-----D-----K 464
Db 4440 VISWMRDSNIAIQNNEKYQILGTLTALRNVLPPDDG-FYHCIAKSDAGQKIATRKLIYNK 4498
QY 465 -----C-----VHI-----W-----N----- 470
Db 4499 PSDRPAPIWVEDEKPKKTEYIMIDRGDTPDDNPQLLPKQVEDSSLSNGSIAYRCMPG 4558
QY 471 -----T-----Q-----T-----CA-----LV----- 477
Db 4559 RSRVTLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPHPTITWAKDGKLIEDSKFEIA 4618
QY 478 --H-----S----- 479
Db 4619 YSHLKYTLNSTSDSGEYTCMAQNSVGSSTVSAPINVNDNILPTPKSSNQKNVAVITCYE 4678
QY 480 -----Y-RG-T--G-----GI----- 486
Db 4679 RNQAYSRLGTWEYGVMPKPNLAGIHFMMNGSLIVLDTSSLKEGDLLEYTCVKRNRRHS 4738
QY 487 -----FE-----V-C-----W-N-----A----- 493
Db 4739 IPHLTSAFEGVPEVKTIDKVVNNGSDSVLDCEVTSDDLTHVWTKNDQKMLDDAIYV 4798
QY 494 -----A-----G-----D-----KV--G-----AS----- 501

Db 4799 LPNNSIVLLNVEKYDEGVYKCVASNSIGKAFDDTQLNVYGGSSRRREAYKKENEDASTTTI 4858
QY 502 -----A-----S--DG----- 505
Db 4859 TTTSPTTTTTETPLTTIIPALITLPAKOYPTDDVHEGSANDDGFPTTQDSLFEFNPPL 4918
QY 506 --SV--CV-----LD----- 511
Db 4919 HPEISVWNTDCAGTINENGDCVDKQKTHNLKILTGENHCPEGFAMNPHPRICEDLDECA 4978
QY 512 -----L-----R-----K 514
Db 4979 FYQPCDFECINVDGFGQCNCPGLGYELAERGCGRDVNECESVRCEGDK 5024

RESULT 13
US-10-369-493-6861
; Sequence 6861, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6861
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6861

Query Match 71.0%; Score 2596.6; DB 14; Length 5175;
Best Local Similarity 9.0%; Pred. No. 2.1e-29;
Matches 454; Conservative 48; Mismatches 10; Indels 4514; Gaps 390;

QY 1 M-----SI-----SS-----D-----E----- 7
Db 1 MGRSPSWLYGVLLGILLATTCSSVNDKNDPTGKSSLAFFDITGSMFDDLVQVREGAAK 60
QY 8 -----V-----N-----F--L--VY----- 13
Db 61 IFKTVMAQREKLIYNYVPHDPYLGELIINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120
QY 14 -----R--Y-----L--OE--SG--F-----SHSAP-T 27
Db 121 ILKALQISLPSSFIYVFTDARSQVHLEDEVLTIQEQSSVVFVMTGCGNRTHPGFRT 180
QY 28 -----FG-I---KS-----H-ISO-----S-NI----- 39
Db 181 YEKIAAASFGQVFHLEKSDSVTVLEVRHAVKQKVHLMYEARERGGTVSRNIPVDKHLS 240
QY 40 -----N-----GA-----LV-----P-P----- 46
Db 241 ELTISLGDKDDNLDIVLRPEGRTVDKRLYSKEGGTIDLKNVCLIRLKDPSFGVWTV 300
QY 47 -----A-----A-----LI--S--I----- 52
Db 301 NTNSRLKHTIRVFGHGAVDYFKYGFASRPLDIRARPRVLNQDTYLLINMTGLIPPGTV 360
QY 53 --I-----Q-----KGL-----Q-Y-----VE-----AE- 63
Db 361 GEIDLVDYHGHSYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYDEBNYEFMRAPT 420
QY 64 -----V-----S-I-----N-----ED-----GTLF----- 73

QY 284 ---AG---V---DK-T---TI---I--- 292
Db 2640 SGRYTCIVRPAGEARKLFDFAVNDPPSISDELSSANIQTIVPPYVEINCVVSGSPHPK 2699
QY 293 ---W---D--A---H---T---G--- 298
Db 2700 VYWLFDKPLEPDSAAVELTNNGETLKIVRSQVEHAGTYTCEAQNNGVKARKDFLVRVTA 2759
QY 299 ---E---AK---Q--- 302
Db 2760 PPHFEKEBEVARVGDMLLTCAESSVPLSSVYVYHHADESQVGVITTSKYAANEKTLN 2819
QY 303 ---Q---F--- 304
Db 2820 VTNIQLDDSGFYCTAVNEAGITKFKLIVETPYFLDQQLYPILGLKRLTLDCSATG 2879
QY 305 P--F--- 306
Db 2880 TPTPTILFMKDGKRLNESDEVDIIGSTLVDPKEVEGRYTCTIAENKAGRSEKDMMEV 2939
QY 307 ---H---S-- 308
Db 2940 LLPPLSKWINVEVOAGDPLTLECPIDTSGVHITWSRQFGKQGLDMRAQSSSKSL 2999
QY 309 ---A-P--A--L---D---V-D--- 315
Db 3000 YIMQATPEDADSVCIAVNDAGAEAVQVYVNTTPPKIFCDSSFSTTEIVADTTLBIPCRT 3059
QY 316 ---W---Q--- 317
Db 3060 EGIPPEISWFLDKGPILEMPGVYTKQGLSLRIDNIKENQEGRYTCVAENKAGRAEQDT 3119
QY 318 ---S---N---T---F--- 322
Db 3120 YVEISPPRVWASEVMRVVEGRQTTIRCEVFGNPEPVVWNLKDGEPYTDLLQFSTKLS 3179
QY 323 ---A---SC---S---TD---M--C- 330
Db 3180 YLHLRETTLDGGTYTCIATNKAGESQTTDVEVLVPPRIEDEERVLOKKEGNTYVWHQ 3239
QY 331 ---I---HV---C--- 335
Db 3240 VTGRPVYVTKRNGKEIEQFNPLHNRATRADEKYSIASNEAGTAVADFLIDVFTK 3299
QY 336 ---L--GO---D---RP---I-K- 343
Db 3300 PTFETHETFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMNDNIIILSPRGDTLMILKA 3359
QY 344 ---T---F---Q--- 346
Db 3360 QRPDGLYTCVATNSYGDSEQDFKNVYTKPYIDETIDQTPKAVAGEIILKCPVLGNPT 3419
QY 347 ---G---HT---N---E---VN---A-I- 355
Db 3420 PTVTKRGDDAVPDSRRTIVANNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479
QY 356 ---K---W---D---P---T- 360
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYLDYRYSISPDGSHITI 3539
QY 361 ---G---N---L---LA---S---C--S 368
Db 3540 NKAKLSGGKYICRASNEAGTSIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPIS 3599
QY 369 ---D---D-M--- 371
Db 3600 GIPQPDVIWTKGMDINWDSRVLQAQNETFGIENVQVTDQGRYTCATNRGKASHDF 3659
QY 372 ---TL---K---I---W---S---M-- 378
Db 3660 SLDVSPPEFDIHGTOPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLD 3719

QY 379 ---K--Q---D--- 381
Db 3720 NVDISDGRKLTISQASLENAGLYTCIALNRAGEASLEPKVEILSPVIDISRNDVQPOV 3779
QY 382 ---N---C-V--H---D---L--Q--Q--H--- 390
Db 3780 AVNQPTIMRCAVTHGPPFSIKWLKNGKEVTDENIRIVEQGVQLILRTDSDHAGKSCV 3839
QY 391 ---N---KE---I--Y--TI---KW--- 399
Db 3840 AENDAGVKELEMLVDVFTPPVVSUKSDNPIKALGETITILFCNASGNPYQLKWAKGSLI 3899
QY 400 ---SP---T--G---P---GT---N--- 407
Db 3900 FDPDGCARISLKGARLDIPLHLKKTVDGTYTQALNAAGTSEASVSDVLVPPPEINRDGID 3959
QY 408 ---N---P--- 409
Db 3960 MSPRLPAQOSLTLOCLAQKFPQMRWTLNGTALTHTSTPGITVADSTFIQNNVSLSDK 4019
QY 410 ---N-A--NLM---LA---S--- 417
Db 4020 GYVTCYAEENVAGSDNLMYNDVVPQAPVISNGGKTQVIEGELAVIECLVBSGYPAPQVSWLR 4079
QY 418 ---A-SFDS--- 422
Db 4080 NGRNVETGVQVRYVTDGRMLTIEARSLSDSGLYLSATNEAGSAQAYTLEVLVSPKII 4139
QY 423 ---T---VR-L--W---DV--- 429
Db 4140 TSTPGVLTSSSGSKFSLPCAARGVDPPIISWTNGNDIKDNGHGTIGADGTLHIEKAE 4199
QY 430 ---D---R---G---I-C-I--- 435
Db 4200 RHLIIECTAKNDAGADTLEFPVQTVAPKISTSGHRYNGSEGETEVIKCEISESESEFS 4259
QY 436 ---H---T-L-- 440
Db 4260 WSKNGVPLLPNNLIFSEDYKLIKILSTRLSDOGEYSCTAANKAGNATOKTNLNVGVAPK 4319
QY 441 ---H--Q---E---P-V-Y---SVAF-- 450
Db 4320 IMERPRTQVHVHKGQDQVTLWCEASGVQPAPITWYKONELLTNGTVDATATTKKSVFSSI 4379
QY 451 SP---D---G--- 454
Db 4380 SPQAGVYTCKAENWVASTEEDIDLIVMIPPEVPMVSTNPROTVFLSCNATGIPEP 4439
QY 455 ---R---Y---LA---SGSF---D---K 464
Db 4440 VISWMDSNIAIQNNEKYQILGTTLAIRNVLPDDG-FYHCIAKSDAGQKIATKRLIVNK 4498
QY 465 ---C---VHI---W---N--- 470
Db 4499 PSDRPAPIWVEDEKPKKTYMIDRGDTPDNPOLLPMKQVDESSLNGSIAYRCMPGP 4558
QY 471 ---T---Q---T--GA---LV--- 477
Db 4559 RSRVTLLHAAPQFIVKPKNTTAAIIVELRCSAAGPPHPTITWAKDKLIEDSKFEETA 4618
QY 478 ---H---S--- 479
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAPINVDNNILPTPKSSNQKNVAVITCYE 4678
QY 480 ---Y-RG-T--G---GI--- 486
Db 4679 RNQAYSRLGTWBYNGVPMKPNLAGIHFMMNGSLIVLDTSSLKEGDLLEYTCVKVNRNRHS 4738
QY 487 ---FE---V-C---W--N---A--- 493
Db 4739 IPHLTSAFEGVEVKTIDKVEVNGSDVLDCEVTSDDLTHVVTKNQKMLDDDAIYV 4798
QY 494 ---A---G---D---KV--G---AS--- 501

Db 1200 CPAKALPPVRLWTEYEGEKIDSLIPHTIREDCALVQNVKLENTGVFVQVSNLAGEDS 1259
QY 153 -----G-DV-----EIP-----P-----NK-AV----- 163
Db 1260 LSYTLTVHEPKIIEVPGVVDVVGFTIBIPCRATGVPVEVIRTNKNGIDLKMDKKFS 1319
QY 164 V-----LR-----G-----H-----E-----S----- 170
Db 1320 VDNLGLTIRIYEADKNDIGNVCVTVNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGDR 1379
QY 171 -----E-----V-----F-----I-----C----- 175
Db 1380 VELKCYEASPPASVTWFRGIAIGTDTKGYVVESDGLTVIQSASVEDATIYTCASNPA 1439
QY 176 -----A----- 176
Db 1440 GKAEANLQVTVIASPDIKDPDVVTOBSIKESHFPFSLYCPVFSNPLPQISWYLNKPLIDD 1499
QY 177 -----W----- 178
Db 1500 KTSWKTSDDKRLHVFKAITDSGVYKCVARNAAAGESKSFQVEVIVPLNLDSEKYYKKV 1559
QY 179 -----PVS----- 181
Db 1560 FAKEGEVTLGCPVSGFPVPQINWVVDGTVEPKKYKATLSNDGLTLHFDSSVVKQEG 1619
QY 182 -----D-----L-----LA----- 186
Db 1620 NYHCVAQSGNILDIDVELSVLAVPIVGEDDNLVFLGKDLSLQLOTESDDKTTFFVWS 1679
QY 187 -----GS-----G-----DS-----T-----A-----R----- 194
Db 1680 INGESDRPDNVQIPSDGHLRYITDAKPNNGKYMCMVTVNSAGKARTLTLVLEPPVVF 1739
QY 195 -----IW-----N----- 199
Db 1740 EPVFEANQKLGNNPILOQVGTGNPKPTVIWKIDGNDVDKSWLPDESLSLLRIEKLTKG 1799
QY 200 -----EN----- 207
Db 1800 SAQISCTAENKAGTASRDPFIQIAAPTAKNEGQDQETIPRESETITLDCPVSIGDFQITW 1859
QY 208 -----Q-----L-----VL-----R-----H-----C----- 216
Db 1860 MKQGLPLTENDAIFLTDNLTILNARDHEDIYTCVANTAGQVSKDFDVVVQVLPKIK 1919
QY 217 -----EG-----G-----Q-----DVP----- 223
Db 1920 NAVVLEINEGEBIILTCDAEGNPTAKWDFNQGLPKEAVFVNNHHTVWVNNVTKYHT 1979
QY 224 -----SNK-----DV----- 232
Db 1980 GYVKYATNKVQAVKTIHVHRTKPRFESGLTESELTVNLTRSITECDVDDAIGVGIS 2039
QY 233 W-N-N-----SE-G-TL-----L-----A-----T-----GSY----- 245
Db 2040 WTVNGKPLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVINEAGVATKTNLFVQVPP 2099
QY 246 -----DG----- 248
Db 2100 TIVNEGGEVTVIENSLVLPCEVTGKPNFVVTWKDGRPVGDLKSVQVLSGQQPKIVHA 2159
QY 249 -----A-----R----- 250
Db 2160 ETIAHGSYICMAKNDVGTAEISFDVDIITRPMIQKGIKNIVTAIKGALPFKCPIDDDKN 2219
QY 251 -----IW----- 257
Db 2220 FKQIILWLRNYQPIDLEADARITLSNDRRLTILNVTEDEGQYSCRVDNAGENSFDF 2279
QY 258 -----L-----A-----S-----TL----- 265

Db 2280 KATVLVPPTIIMLDKKNKTAVEHSTVTLSCPATGKPEPDIITWFKDGEAIHIEIADIIP 2339
QY 266 ----- 265
Db 2340 NGELNGNQLKITRIKEGDAGKYTCADNSAGSVEQDVNNVNIITIPKIEKDIPSDYESQ 2399
QY 266 -----K-G-P----- 269
Db 2400 NERVVISCVPYARPPAKITWLKAGKPLQSDKFVKTSANGQKLYLFKLRETDSSKYTCIAT 2459
QY 270 -----F-----A-----L-----K-----W----- 274
Db 2460 NEAGTDKDRDFKVSMLVAPSFDBFNIVRRITVNSGNPSTLHCPAKGSPSPTITWLKDGNAI 2519
QY 275 -----N-----K----- 276
Db 2520 EPNDRYVFFDAGRQLOISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPVIDGRREA 2579
QY 277 -----K-G-N-----FI-----LSA----- 284
Db 2580 VAVIEGFSSELCDSNSTGVDVEMQKDLGTINODTLRGDSFIOIPSSGKKMSFLSARKSD 2639
QY 285 -----G-----V-----DK-----T-----TI-----I----- 292
Db 2640 SGRYTTCIVRNPAAGEARKLFDFAVNDPPSISDBLSANIQTIVPYYPVEINCVVSGSPHK 2699
QY 293 -----W-----D-----A-----H-----T-----G----- 298
Db 2700 VTWLFDDKPLEPDSAAVELTNNGETLKIVRSQVEHAGTYTCERQNNVVGKARKDFLVRVTA 2759
QY 299 -----E-----AK-----Q----- 302
Db 2760 PPHFEKEREEVAVGDTMLLTCAESSVPLSSVYVHHADESQNGVITSKYAANEKTLN 2819
QY 303 -----Q-----F----- 304
Db 2820 VTNIQLDDEGYCYTAVNEAGITKFKFLIVETPYFLDQKLYPIILKRLTLDCSATG 2879
QY 305 -----P-----F----- 306
Db 2880 TPTPTILFMKQKRLNESDEVDIIGTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVEV 2939
QY 307 -----H-----S----- 308
Db 2940 LLPPKLSKEWINVEVQAGDPLTLECPIEDTSGVHITWSRQFGKQDQOLDMAQSSSKSL 2999
QY 309 -----A-P-A-L-----D-----V-D----- 315
Db 3000 YIMQATPEDADSYSCIADVNDAGBAVFOVTNTPPKIFGDSFSTTEIVADITLPIPCRT 3059
QY 316 -----W-----Q----- 317
Db 3060 EGIPPEISWFLDGPILFEMPGVYKQGLSLRINIKPNQEGRYTCVAENKAGRAEQDT 3119
QY 318 -----S-----N-----N-----T-----F----- 322
Db 3120 YVEISEPPRVWASVEMRVVEGRQTTIRCEVFNPEPVVWMLKDGEPYTDLLQFSTKLS 3179
QY 323 -----A-----SC-----S-----TD-----M-----C----- 330
Db 3180 YLHLRETTILADGGTYTCIATNKAGESQTTTDEVLVPPRIEDEERVLQKEGNTYMHQC 3239
QY 331 -----I-----HV-----C-----K----- 335
Db 3240 VTGRPVVYVTVWRNKKEIQFNPVLHIRNATRADEGKYSIASNEAGTAVADFLIDVFTK 3299
QY 336 -----L-----GO-----D-----RP-----I-K----- 343
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGGPFNMDNIIILSPRGDTMLKA 3359
QY 344 -----T-----F-----Q----- 346
Db 3360 QRFDDGLYTCVATNSYGDSEQDFKVNVTKPYIDETIDQTPKAVAGGEIILKCPVLGNPT 3419

QY 347 ---G---HT---N---E---VN---A---I---355
Db 3420 PTVTKRGDDAVPNDSDRHTIINNVDLKINSVTTEDAGQYSCIAYNEAGNLTHYAAEVIG 3479
QY 356 ---K---K---W---D---P---T---360
Db 3480 KPTVRKGNLYEVIENDTITWDCGVTSRPLPSISWFRGDKPVVLYDRYSISPDGSHITI 3539
QY 361 ---G---N---L---LA---S---C---S---368
Db 3540 NKAKLSGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIVLECPIS 3599
QY 369 ---D---M---D---D---D---D---371
Db 3600 GIPODVIWTKNGMDINWTSRVLAAQNNTFFGIENQVQVTDQGRYTCATNRGKASHDF 3659
QY 372 ---TL---K---I---W---S---M---378
Db 3660 SLDVLPPEFDIHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLDLD 3719
QY 379 ---K---Q---D---D---D---D---381
Db 3720 NVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPVVIDISRNDVQPV 3779
QY 382 ---N---C---V---H---D---L---Q---Q---H---390
Db 3780 AVNQPTIMRCVAGTGHPPFSIKWLKNGKEVTDENIRIVEGQVQLIILRTSDHAGKWSVCV 3839
QY 391 ---N---KB---I---Y---TI---KW---399
Db 3840 AENDAGVKELEWLDVETPPVSVKSDNPIKALGETITLFCNAGSNPYQPKWAKGSLI 3899
QY 400 ---SP---T---G---P---GT---N---407
Db 3900 FDSPDGARISLKGARLDIPLHKTDVGYTCQALNAAAGTSEASVSDVLVPPPEINRDGID 3959
QY 408 ---N---N---P---P---P---P---409
Db 3960 MSPRLPAQOSITLQCLAQKPVQPMRWLTNGALTHTPGITVASDSTFIQINNVSLSDK 4019
QY 410 ---N---A---NLM---LA---S---417
Db 4020 GVTCTAENAVGSDNLMVNDVQAPVISNGTKQVIEGELAVIECLVEGYPAPQVSWLR 4079
QY 418 ---A---SPDS---A---SPDS---422
Db 4080 NGRVETGVQGVRYVTDGRMLTIEARSLDSGIYVLSATNEAGSAQOAYTLEVLVSPKII 4139
QY 423 ---T---VR---L---W---DV---DV---429
Db 4140 TSTPGVLTPSSGSKFSLPCAARGVPDPFIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4199
QY 430 ---D---D---R---G---I---C---I---435
Db 4200 RHLIYECTAKNADAGADTLEFPVQTVIAPKISTSGNRYINGSEGTETVKEICEISESEFS 4259
QY 436 ---H---H---T---L---T---T---K---440
Db 4260 WSKNGVPLPSNNLIFSEDYKLIKILSTRLSDOGEYSCTAANKAGNATQKTNLNVGVA PK 4319
QY 441 ---H---Q---E---P---V---Y---SVAF---450
Db 4320 IMERPRTQVHVKGQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTKKSVIFS 4379
QY 451 SP---D---D---D---D---G---454
Db 4380 SPSQAGYTTCAENWVASTEEDIDLIMIPPEVPERMVSTNPRTOTVFLSCNATGIPEP 4439
QY 455 ---R---Y---LA---SGSF---D---K---464
Db 4440 VISWMDSNIAIQNNEKYQILGTLTALRNVLPPDDG-FYHCIAKSDAGQKIATRLKLVNK 4498

QY 465 ---C---VHI---W---N---470
Db 4499 PSDRPAPIWCEDEKGPKTEYMDIDRGDTPDDNPQLPMKQVEDSSLANGSIAYRCMPGP 4558
QY 471 ---T---O---T---GALV---477
Db 4559 RSRTVLLHAPOFIVKPKNTTAAIGAIVELRCSAAGPPHPTITWAKDGKLIEDSKFEIA 4618
QY 478 ---H---S---S---479
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAVITCYE 4678
QY 480 ---Y---RG---T---G---GI---486
Db 4679 RNOAYSRGLTWEYNGVMPKRLAGHFMNNGSLVILDTSLKEGDLLEYTCVKVRNRHRS 4738
QY 487 ---FE---V---C---W---491
Db 4739 IPHLTSAFEGVPEVKTIIDKVEVNGSDSVLDCVTSDDLTHVVTWKNDQKMLDDDAIYV 4798
QY 492 ---N---A---A---GD---K---V---498
Db 4799 LPNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQLNVYEGDFPLTGFEGSGINIDSS 4858
QY 499 ---G---AS---A---A---502
Db 4859 NAGGSRREAYKKNEDASTTTITTSPTTTTETPLTTIIPALITLPAKQVPTDDYHE 4918
QY 503 ---S---DG---SV---CV---509
Db 4919 GSANDDGFPTQDLSLFEENPPLHPEISVVNTDCAGTINENGDCVDKDGKTHNLKLTGE 4978
QY 510 ---LD---LD---L---L---512
Db 4979 NHCPEGFAMNPHTRICEDLDECAFYQPCDFECINYDGGFCQCPLGYELAEBCRDVNEC 5038
QY 513 ---R---K---K---514
Db 5039 ESVRCEDGK 5047

RESULT 15

US-10-369-493-6858
; Sequence 6858, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6858
; LENGTH: 5198
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6858

Query Match 70.9%; Score 2595.3; DB 14; Length 5198;

Best Local Similarity 9.0%; Pred. No. 2.3e-29;

Matches 453; Conservative 50; Mismatches 9; Indels 4537; Gaps 389;

QY 1 M-----SI-----SS-----D-----E-----7

Db 1 MGRSPSWLYGVLLGILLALLATTCSSVNDKNDPTGKSSLAFFVFDITGSMFDDLVQVREGAAK 60

Qy 8 ---V---N---F---L---VY---13
Db 61 IFKTVMAQREKLIYNYIMVDFHPYLGEIINTTDSYFMRQLSKVYVHGSGDCPEKTLTG 120
Qy 14 ---R---Y---L---QE---SG---F---SHSAF-T 27
Db 121 ILKALQISLPSFIYFTDARSXDYHLEDEVLNTIOEKQSSVVFVMTGCGNRTHPGFR 180
Qy 28 ---FG-I---KS---H-ISO---S-NI---39
Db 181 YEKIAAASFGQVHLEKSDYSTVLEYVRHAKQKVHLMYEARERGTVSRNIPVDKHL 240
Qy 40 ---N---N---GA---LV---P-P---46
Db 241 ELTISLGDKDDNLDIVLRDPGRVTDKRLYSKEGGTIDLKNVKILRLKDPSPGWTV 300
Qy 47 ---A---A---A---LI---S-I---52
Db 301 NTSRLKHTIRVFGHGAUDFKYGFASRPLDRIELARPRVLNQDTYLLINMTGLIPPGTV 360
Qy 53 ---I---Q---KGL---O-Y---VE---AE-63
Db 361 GEIDLVDYHLSYKAVASPHRTNPNMYFAGPPVPPKGLFFVRVQGVDEDNIEFMRIAPT 420
Qy 64 ---V---S-I---N---ED---GTLF---73
Db 421 AIGSVIVGGPRAPMSPHQEFVGRDLNLSCTVESASAYTIYWKGTGEDIIIGPLFYHNTD 480
Qy 74 ---D---L---G---R-P-I---E---79
Db 481 TSVMTIPELSLKDAGEYECRVISNNGNYSVKTRVETRESPEIFGVRNVSVPLGEAFLH 540
Qy 80 ---S---L---SL-I---DA---86
Db 541 CSTRSAGEVEIRTRGATVNGPNTERNPTNGTLKIHVTRADAGVYECMARNAGMST 600
Qy 87 ---VM---P-DV---V---92
Db 601 RKMRLDIMEPPSVKVTPODVYFNMREGVNLSCAMGDPKPEVHVFYKGRHLLNDYKYVG 660
Qy 93 Q---T---R---Q-QAYRD---KLA---Q---105
Db 661 QDSKFLYIRDATHHDEGTVECRAMSAQGA-RDITDMLATPPKVEIIONKMMVGRGDRV 719
Qy 106 ---Q---Q---Q---A---A---A---111
Db 720 SFECKTIRGPKPKIRWFKNKGLIKPDDYIKINEQLHMGAKDEADAGAYSCVGENMAG 779
Qy 112 ---A---A---A---A---114
Db 780 KDQVANLSVGRVPTIIESPHTVRVNIERQVTQLCLAVGIPPEPEIEWQKGNVLLATLNP 839
Qy 115 ---A---A---A---ASQ---120
Db 840 RYTQADGNLLITDAQIEDQOFTCIARNTYGOOSQSTILMTGLVSPVLGHVPPBEQLI 899
Qy 121 ---Q---Q---Q---GS---123
Db 900 EGQDLTSCVVVLGTPKPSIVWIKDKPVEEGPTIKIEGGSLRLRGNPNKDEGKYTCI 959
Qy 124 ---A---A---K---N---GE---128
Db 960 AVSPAGNSTLHINVQLIKKPEFVYKPEGGIVKPTISGMDEKHVAVNTHDVLDEGFA 1019
Qy 129 ---NT---A-N-G-133
Db 1020 IPCVSGTPTPIITWYLDGRPIFNSRDFVTADNTLIVRKADKSYSGVYTCOATNSAGD 1079
Qy 134 -E- ---E-N-136
Db 1080 NEQKTTIRIMTNPISFGOSSFNMVDDLFTIPCDVYGPKPVTWLLDDKPTGEGVNE 1139
Qy 137 -G- ---AH---TI-A-N---N---144

Db 1140 DGSLLTIPNVNEAHRGFTTCHQNAAGNDTRTTLTVHTTPTINAENQEKIALQNDIDVLE 1199
Qy 145 ---HT---D---MME---V---D-152
Db 1200 CPAKALPPPPVRLWYEGEKIDSOLIPHTITREDGALVQNQKLENTGVFVCOVSNLAGEDS 1259
Qy 153 ---G-DV---EIP---P---NK-AV-163
Db 1260 LSYTLTVHEKPKIISVPGVWDVVKGTIIEIPCRATGVPVIRTKNGKIDLKMDEKKFS 1319
Qy 164 V---LR---G---H---E---S---170
Db 1320 VDNLGLTRIYEADKNDIGNVNCVVTNEAGTSQMTTHVDVQEPPIILLPSTOTNNTAVVGDR 1379
Qy 171 ---E---V-F---I-C---175
Db 1380 VELKCYVEASPPASVTWFRRGIAIGTDTKGYVVESDGTLVIOSASVEDATITCKASNPA 1439
Qy 176 ---A---A---176
Db 1440 GKAEANLQVTVIASPDIKDPDVVTQESIKESHPPSLYCPVFSNPLPQISWYLNKPLIDD 1499
Qy 177 ---W---N---178
Db 1500 KTSWKTSDDKRKLHVFKAKITDSGVYKCVARNAAGEGSKSFQVEVIVPLNLDSEYKVKV 1559
Qy 179 ---PVS---181
Db 1560 FAKGSEVTLGCPVSGFPVQINWVVDGTVEFGKYKYGATLSNDGLTLHFDSSVKQEG 1619
Qy 182 ---D-L-LA---S-186
Db 1620 NYHCVAQSKENILDIDVELSVLAVPIVGEDDNLVFLGKDISLSCDLQTESDDKTTFVWS 1679
Qy 187 ---GS---G---DS---T---A-R---194
Db 1680 INGSERSDPNVQIPSDGHRLYITDAKPENNGKMYCRVTNSAGKABERTLTLDVLEPPFV 1739
Qy 195 ---N---N---LS---199
Db 1740 EPVFEANQKLGNNPIILOCQVTGNPKPTVIWKIDGNDVDKSWLFDSELSLRIEKLTKG 1799
Qy 200 ---EN---S---S---T---S-GS---T-207
Db 1800 SAQISCTAENKAGTASRDPFIQNIAPTFKNEGDQETIFRESEITILDCPVSIGDQITW 1859
Qy 208 ---Q---L-VL---R-H---C---IR-216
Db 1860 MKQGLPLTENDAIFTLDNTRLTILNANRDHEDIYTCVANNTAGOVSKDFDVVVQVLPKIK 1919
Qy 217 ---EG---G---Q-DVP---223
Db 1920 NAVVTLEINEGEBEILITCDAEGNPTPTAKWDFNQGLDLPKEAVFVNNHTVVNNVTKYHT 1979
Qy 224 ---SNK---DV---T-S-L-L-D-232
Db 1980 GUYKCVATNKVQAVKTIINVHVTRKPRFESGLTESELVNLTRSTILECDVDDAIGVGIS 2039
Qy 233 W-N---SE-G-TL---A-T---GSY---245
Db 2040 WTVNGKFLAETDGVQTLAGRFLHIVSAKTDHGSYACTVTNEAGVATKTNLFVQVPP 2099
Qy 246 ---DG---F---248
Db 2100 TIVNEGTYVIENNSLVLPCBVTGKPNPVVTVTKDGRVGLKSVQVLSGQQPKIVHA 2159
Qy 249 ---A---R---250
Db 2160 EIAHKSYICMAKNDVGTAEISFDVDIITRPMIOKIKNIIVTAIKGALPFPKPIDDDKN 2219
Qy 251 ---IW---T---K-D-G-N---257

Db 2220 FXGQIILWLNRYQPIDLEAEDARITRLSNDRLTILNVTEDEGQYSCRVKNDAGENSDFD 2279
Qy 258 -----L-----A--S--TL-----GQ--H----- 265
Db 2280 KATVLVPPTIIMLDKDKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAIHENIADIIP 2339
Qy 266 -----G-----HT-----N--E-----VN-----A--I-- 355
Db 3420 PTVTWKRGDDAVPNDSTRHTIVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVI 3479
Qy 356 -----K-----G-----N-----L-----LA-----S--C--S 368
Db 3480 KPTFVRKGNLYEVIENDITMDGVTSRPLPSISWFRGDKPVYLDVRSISPDGSHITI 3539
Qy 361 -----G-----N-----L-----LA-----S--C--S 368
Db 3540 NKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPIS 3599
Qy 369 -----D-----D--M-----D-----D-----D----- 371
Db 3600 GIPQDVITWKGMDINMTDSRVILQONNETFGIENVQVTDQGRYTCTATNRGGKASHDF 3659
Qy 372 -----TL-----K-----I-----W-----S-----M-- 378
Db 3660 SLDLVSPBFDIHGTPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKDSRALDGLD 3719
Qy 379 -----K-----Q-----I--Y--TI-----KW----- 399
Db 3720 NVDISDDGRKLITISQASLENAGLYTCIALNRAGEASLEFKVEILSPPIDISRNDVQPV 3779
Qy 382 -----N-----C-V--H-----D--L--O--O--H----- 390
Db 3780 AVNQPTIMECAVTHGPFPSIKWLNKNGKEVTDENIRIVEQGVQLLRTDSDHAGKSCV 3839
Qy 391 -----N-----KE-----I--Y--TI-----KW----- 399
Db 3840 AENDAGKELEMLVDVFTPPVSVKSDNPIKALGETITLFCNAGSNPYPOLKWKAGGSLI 3899
Qy 400 -----SP-----T--G-----P-----GT-----N----- 407
Db 3900 FDSPDGARISLKGARLDIPHLKKTVDGYTCQALNAGTSEASVSVDVLVPPPEINRDGID 3959
Qy 408 -----N-----N-----P-----P----- 409
Db 3960 MSPRLPAQSLTLQCLAQKPVQMRWTLNGTALHTSTFGITVADSTFIQINNVSLSK 4019
Qy 410 -----N--A--NLM-----LA-----S----- 417
Db 4020 GYVTCYAENVAGSDNLMYNDVVDVQAPVISNGGKQVIEGELAVIECLVEGYPAPQVWMLR 4079
Qy 418 -----A--SFDS-----A----- 422
Db 4080 NGRVETGVQVRYVTDGRMLTIIIEARSLDSGIYLCSTATNEAGSAQAYTLEVLVSPKII 4139
Qy 423 -----T-----VR--L-----W-----DV----- 429
Db 4140 TSTPGVLTSPSSGSKFSLPCAVERGYPDPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4199
Qy 430 -----D-----D-----R-----G-----I--C--I----- 435
Db 4200 RHLIYECTAKNDAGADTLEFPVQTVIAPKISTSGNRYINGSEGTETVIKCEIESESEFS 4259
Qy 436 -----H-----T--L-----K----- 440
Db 4260 WSKNGVPLLPNNLIFSEDYKLIKLSTRLSDQGEYSCTAANKAGNATQKTNLVGVAPK 4319
Qy 441 -----H--Q-----E-----P--V--Y-----SVAF----- 450
Db 4320 IMERPRTQVHKGDQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTKKSVIFSSI 4379
Qy 451 SP-----D-----D-----G----- 454
Db 4380 SPSQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMNVSTNPRQTVFLSCNATGIP 4439

Db 2220 FXGQIILWLNRYQPIDLEAEDARITRLSNDRLTILNVTEDEGQYSCRVKNDAGENSDFD 2279
Qy 258 -----L-----A--S--TL-----GQ--H----- 265
Db 2280 KATVLVPPTIIMLDKDKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAIHENIADIIP 2339
Qy 266 -----G-----HT-----N--E-----VN-----A--I-- 355
Db 3420 PTVTWKRGDDAVPNDSTRHTIVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVI 3479
Qy 356 -----K-----G-----N-----L-----LA-----S--C--S 368
Db 3480 KPTFVRKGNLYEVIENDITMDGVTSRPLPSISWFRGDKPVYLDVRSISPDGSHITI 3539
Qy 361 -----G-----N-----L-----LA-----S--C--S 368
Db 3540 NKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPIS 3599
Qy 369 -----D-----D--M-----D-----D-----D----- 371
Db 3600 GIPQDVITWKGMDINMTDSRVILQONNETFGIENVQVTDQGRYTCTATNRGGKASHDF 3659
Qy 372 -----TL-----K-----I-----W-----S-----M-- 378
Db 3660 SLDLVSPBFDIHGTPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKDSRALDGLD 3719
Qy 379 -----K-----Q-----I--Y--TI-----KW----- 399
Db 3720 NVDISDDGRKLITISQASLENAGLYTCIALNRAGEASLEFKVEILSPPIDISRNDVQPV 3779
Qy 382 -----N-----C-V--H-----D--L--O--O--H----- 390
Db 3780 AVNQPTIMECAVTHGPFPSIKWLNKNGKEVTDENIRIVEQGVQLLRTDSDHAGKSCV 3839
Qy 391 -----N-----KE-----I--Y--TI-----KW----- 399
Db 3840 AENDAGKELEMLVDVFTPPVSVKSDNPIKALGETITLFCNAGSNPYPOLKWKAGGSLI 3899
Qy 400 -----SP-----T--G-----P-----GT-----N----- 407
Db 3900 FDSPDGARISLKGARLDIPHLKKTVDGYTCQALNAGTSEASVSVDVLVPPPEINRDGID 3959
Qy 408 -----N-----N-----P-----P----- 409
Db 3960 MSPRLPAQSLTLQCLAQKPVQMRWTLNGTALHTSTFGITVADSTFIQINNVSLSK 4019
Qy 410 -----N--A--NLM-----LA-----S----- 417
Db 4020 GYVTCYAENVAGSDNLMYNDVVDVQAPVISNGGKQVIEGELAVIECLVEGYPAPQVWMLR 4079
Qy 418 -----A--SFDS-----A----- 422
Db 4080 NGRVETGVQVRYVTDGRMLTIIIEARSLDSGIYLCSTATNEAGSAQAYTLEVLVSPKII 4139
Qy 423 -----T-----VR--L-----W-----DV----- 429
Db 4140 TSTPGVLTSPSSGSKFSLPCAVERGYPDPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4199
Qy 430 -----D-----D-----R-----G-----I--C--I----- 435
Db 4200 RHLIYECTAKNDAGADTLEFPVQTVIAPKISTSGNRYINGSEGTETVIKCEIESESEFS 4259
Qy 436 -----H-----T--L-----K----- 440
Db 4260 WSKNGVPLLPNNLIFSEDYKLIKLSTRLSDQGEYSCTAANKAGNATQKTNLVGVAPK 4319
Qy 441 -----H--Q-----E-----P--V--Y-----SVAF----- 450
Db 4320 IMERPRTQVHKGDQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTKKSVIFSSI 4379
Qy 451 SP-----D-----D-----G----- 454
Db 4380 SPSQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMNVSTNPRQTVFLSCNATGIP 4439

QY 455 -----R-----Y-----LA-----SGSF-----D-----K 464
Db 4440 VISWEDSNIAIONNEKYQILGTTLAIRNVLPPDDG-FYHCIAKSDAGQKIATRKLIYNK 4498
QY 465 -----C-----VHI-----W-----N-----470
Db 4499 PSRPAPIWVECKEKGPKKTEYMDIRGTPDDNPQLLPWKDVEDSSLNGSIAYRCMPGP 4558
QY 471 -----T-----Q-----CALV-----477
Db 4559 RSRRTVLLHAAQPIVVKPKNTTAAIGAIIVELRCSAAGPPHPPTITWAKDGKLIEDSKFEIA 4618
QY 478 --H-----S-----479
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAPINVDNNILPTPKPSSNQKNVAVITCYE 4678
QY 480 ----Y-RG-T-----G-----486
Db 4679 RNQAYSRLTWEYNGVMPKPNLAGIHFMMNGSLVILDTSSLKEGDLLEYTCVKNRRRHS 4738
QY 487 -----FE-----V-C-----W-----491
Db 4739 IPHLISAFEGVPEVKTIIDKVEVNGDSVVLDCVTSDDLTHVVTKNQKMLDDDAIYV 4798
QY 492 ----N-----A-----GD-----K-----V-----498
Db 4799 LPNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQLNVYEGDFLPLTGFECSGINDSS 4858
QY 499 ---G-----AS-----A-----502
Db 4859 NAGSSRRRAYKENEDASTTTTTSPTTTTETPLTTIIPALITLPAKQYPTDDYHE 4918
QY 503 -S---DG-----SV-----CV-----509
Db 4919 GSANDGFGFTQDSLFEFNPPLHPEISVUNTDCAGTINENGDCVDKDKTHNLKILTGE 4978
QY 510 -----LD-----L-----512
Db 4979 NHCPEGFAMNPHTRICEDLDECAFYQPCDFECINYDGFQCNCPLGLYELAEBCRDVNEC 5038
QY 513 ---R-----K 514
Db 5039 ESVRCEDGK 5047

Search completed: January 3, 2005, 15:47:16
Job time : 136 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:09:41 ; Search time 21.6667 Seconds
(without alignments)
2282.558 Million cell updates/sec

Title: US-09-987-701-2
Perfect score: 3659
Sequence: 1 MSISSDEVNFLVRYLQESG.....GDKVCASGDGVCVLDLRK 514

Scoring table: BLOSUM30
Gapop 1.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2596.6	71.0	5175	T20592	hypothetical prote
2	2395.3	70.9	5198	T43290	hemacentin precurs
3	2594.7	70.9	7962	I38346	elastic titin - hu
4	2585.4	70.7	26926	I38344	titin, cardiac mus
5	2583.7	70.6	10797	T30192	probable peptide s
6	2580.5	70.5	4544	S02392	alpha-2-macroglobu
7	2580.2	70.5	6658	T13331	projectin - fruit
8	2578.7	70.5	7829	T15789	hypothetical prote
9	2574.9	70.4	4660	T42737	gp330 protein prec
10	2572.2	70.3	6831	A88952	protein unc-22 [im
11	2572.2	70.3	6839	S57242	twitichin [similar
12	2572.2	70.3	7160	T27935	hypothetical prote
13	2571.1	70.3	5825	T12117	polyprotein - fava
14	2569.5	70.2	4545	S25111	alpha-2-macroglobu
15	2568.8	70.2	3461	S58870	reelin precursor -
16	2567.9	70.2	5232	A45086	HC-toxin synthetas
17	2561.3	70.0	4543	A53102	alpha-2-macroglobu
18	2560.7	70.0	5376	T42215	zonadhesin - mouse
19	2559.8	70.0	6805	S20301	titin - rabbit (fr
20	2556.6	69.9	8243	T31307	type I fatty acid
21	2554.4	69.8	4753	A47437	LDL-receptor-relat
22	2543.3	69.5	6669	S55024	nebulin, skeletal
23	2539.8	69.4	15281	S41309	cyclosporin synthe
24	2539	69.4	4767	T31345	hypothetical prote
25	2537.7	69.4	4836	T14346	herc2 protein - mo
26	2537.5	69.3	9376	T14593	syngingomycin synth
27	2536.4	69.3	4302	A38971	polycystic kidney
28	2530.8	69.2	3623	T09456	intrinsic factor-B
29	2529	69.1	6642	T29757	protein UNC-89 - C

30 2526.3 69.0 4861 2 S71752 giant protein p619
31 2526.1 69.0 3623 2 T08618 intrinsic factor-B
32 2522.8 68.9 5138 2 B96695 hypothetcal prote
33 2520.1 68.9 4930 2 E69679 polyketide synthet
34 2519.1 68.8 4464 2 D87755 protein T21E12.4
35 2519 68.8 4639 1 A54794 dynein heavy chain
36 2514.7 68.7 4644 1 A38905 dynein heavy chain
37 2512.8 68.7 4568 2 T08030 dynein heavy chain
38 2512 68.6 4344 1 A53489 dynein heavy chain
39 2509.4 68.6 4572 2 S57908 hypothetcal 527K
40 2508.6 68.6 5369 2 T44807 mycosubtilin synth
41 2504.6 68.5 4447 2 A69679 polyketide contain
42 2504.2 68.4 4845 2 T31067 Bir repeat contain
43 2502.3 68.4 3871 2 T22812 hypothetcal prote
44 2502.1 68.4 3856 2 T51174 ataxia-telangiecta
45 2497.7 68.3 4391 2 A38096 perlecan precursor

ALIGNMENTS

RESULT 1

T20992 hypothetical protein F15G9.4a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T20992; T24733

R:Sulston, J.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z19355

A:Accession: T20992

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5175 <WIL>

A:Cross-references: UNIPROT:Q810L3; EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a

A:Experimental source: clone F15G9

R:Kershaw, J.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z19929

A:Accession: T24733

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5175 <W12>

A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a

A:Experimental source: clone T09B9

C:Genetics:

A:Gene: CESP:F15G9.4a

A:Map position: X

A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3; 1251/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/1

Query Match 71.0%; Score 2596.6; DB 2; Length 5175;

Best Local Similarity 9.0%; Pred. No. 6.6e-53;

Matches 454; Conservative 48; Mismatches 10; Indels 4514; Gaps 390;

QY	1	M-----SI-----SS-----D-----E-----	7
Db	1	MGRSPSLVGLGLLLLATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLVVVRGAAK	60
QY	8	-----V-----N-----F-----L-----VY-----	13
Db	61	IFKTVMAQREKLIYVIMVPHDPYLGEIINTDSTYFMRLSKVYVHGGDCPEKTLTG	120
QY	14	-----R-----Y-----L-----QE-----SG-----F-----SHSAF-T	27
Db	121	ILKALQISLPSSFIYVFTDARSKDVLHLEVLNTIQEKSSVVFVMTGCGNRTHRGFRT	180
QY	28	-----FG-I-----KS-----H-ISO-----S-NI-----	39
Db	181	YEKIAAASFGQVFHLEKSDVSTLVLYRHAHVQKKVHLYMEARERGVTSRNVDPDKHLS	240
QY	40	-----N-----GA-----LV-----P-P-----	46

Db	241	ELTISLGDKDDSNLDIVLRDPEGRVTDKRLYSKEGGTIDLNKXKLIRLKDPSPGWTV	300	Db	1320	VONLGLRIYEADKNDIGNYNCVVTNEAGTSQMTTHVDVQEPPIILPSTQNTNNAVVDGR	1379
Qy	47	-----A-----A-----LI-S-I-----	52	Qy	171	-----E-----V-F-----I-C-----	175
Db	301	NTNSRLKHTIRVFHGGAVDKYGFASRPLDRIBLARPRPVLNQDTYLLINMTGLIPPGTV	360	Db	1380	VELKCVSEASPASVTVFRRGIAIGTDTKGYVVESDGLVIOASVEDATITYTKASNP	1439
Qy	53	-----Q-----KGL-----O-Y-----VB-----AE-63		Qy	176	-----A-----	176
Db	361	GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPPVPPKGLFFVRVQGVDEDNIEFMRIAPT	420	Db	1440	GKAEANLQVTVIASPDIKDPDVVTQESIKESHPPFSLYCPVFSNPLQISWYLNKPLIDD	1499
Qy	64	-----V-----S-I-----N-----ED-----GTLF-----73		Qy	177	-----W-----N-----	178
Db	421	AIGSVIUGGPRAFMSPHQEFVRDNLNLSCTVESASAYTIYWKVTGEDIIGGLFLFYHNTD	480	Db	1500	KTSWKTSDDKRXLHVFKAKITDSGVYKCVARNAAGBSKSFQVEVIVPLNLDSEKYYKV	1559
Qy	74	-----D-----G-----R-----P-I-----E-----79		Qy	179	-----PVS-----	181
Db	481	TSVMTIPELSKDAGEYECRIVSNNGNYSVKTRVETRESPEIFGVRNVSVPLGEAAFLH	540	Db	1560	FAKEGEVTLGCPVSGFPVQINWVVDGTVVBPFGKYKATILSNDGLTLHFDSSVKQEG	1619
Qy	80	-----S-----L-I-----DA-----86		Qy	182	-----D-L-LA-----S 186	
Db	541	CSTRSAGEVEIRWTRGATVNGPNTERNPTNGTLKIHVTRADAGYECMARNAGMST	600	Db	1620	NHCVQAQSKGNILDDIVELSVLAVPIVGEDDNLLEVLGKDISLSCDLQTESDDKTTFVMS	1679
Qy	87	-----VM-----P-DV-----V-92		Qy	187	-----GS-----G-----DS-----T-----A-R-----194	
Db	601	RKRLDIMEPPSVKVTPODYVFNMBGVNLSCEAMGDKPEVHWYFKGRHLLNDYKYQVG	660	Db	1680	INGSSEDRPDNVQI1PSDGHRLYITDAKPENNNGKYMCRVTNSAGKARTILTLDLVLEPPVFV	1739
Qy	93	-----T-----R-----Q-----QAYRD-----KLA-----Q-----105		Qy	195	-----IW-----N-----LS-----199	
Db	661	QDSKFLYIRDATHDEGTVECRAMSQAQ-RDITDMLATPPKVBIIQNMWVGRDRV	719	Db	1740	EPVFEANQKIGNNPILOQVGTGNPKPTVIWKIDGNDVDKSWLFDSELSLRLEKLTGK	1799
Qy	106	-----Q-----Q-----Q-----A-----A-----A-----111		Qy	200	-----EN-----S-----T-----S-----GS-----T-----207	
Db	720	SPECKTIRGPHKIRWFKNKGLIKPDDYIKINEQOLHMGAKDEDAIYSCVGENMAG	779	Db	1800	SAQISCTAENKAGTASRDEFFIQNIAAPTFFKNEGDQETIIPRESEITITLDCPVSLGDFQITW	1859
Qy	112	-----A-----A-----A-----A-----A-----114		Qy	208	-----Q-----L-VL-----R-H-----C-----IR 216	
Db	780	KDQVANLSVGRVPTIIESPHTVRVNIQROVTLQCLAVGIPPEIEWQKGNVLLATLNP	839	Db	1860	MQQGLPLTENDAIFTLDNTRLTILNANRDHEDITYTCVANNTAGQVSKDFVWVQVLPKIK	1919
Qy	115	-----A-----A-----A-----ASQ-----120		Qy	217	-----EG-----G-----O-DVP-----223	
Db	840	RYTLADGNLLITDAQIEDQOFTCIARNTYGOQSOSTTLMVTLGVSPLGHVPPBEQLI	899	Db	1920	NAVVTLEINEGEBEILTCDAEGNPPTAKWDFNQGLPKEAVFVNNHTVNNVTKYHT	1979
Qy	121	-----Q-----GS-----123		Qy	224	-----SNK-----DV-----T-S-L-----D-----232	
Db	900	EGQDLFCVVVLGTPKPSIWIKDDKPVEBEGTIKIEGGSLRLRGNPKDEGKYTCI	959	Db	1980	GUYKCATNKVQAVKTIHVHVKRPFESGLTESELTIVNLTFSTILECDVDDAIGVGIS	2039
Qy	124	-----A-----K-----N-----GE-----128		Qy	233	W-N-----SE-G-TL-----L-----A-T-----GSY-----245	
Db	960	AVSPAGNSTLHINVQLIKKPEFYVKKPEGGIVFKPTISGMDKHAHVAVNSTHVDLDGEGA	1019	Db	2040	WTWNGKFLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTNFLVQVPP	2099
Qy	129	-----NT-----A-N-G-133		Qy	246	-----DG-----F-----248	
Db	1020	IPCWSTGTPPIITWYLDGRIPITPNSRDFVTADNTLIVRKADKSYGVVTCOATNSAGD	1079	Db	2100	TIVNEGGEVTVIENNSLVLPCVETGKPNPVTVTKDGRPVGDLKSVQVLSGGOQFKIVHA	2159
Qy	134	E-----E-N-136		Qy	249	-----A-----R-----250	
Db	1080	NEQKTIIRIMTWMISPGQSSFNWVDDLETPICDVYGDPKPVITMLDDKPTGVVNE	1139	Db	2160	ETAHKSYICMAKNDVGTAEISFDVDIITRPMIQKNIIVTAIKGALPFKPCPIDDDKN	2219
Qy	137	G-----AH-----TI-A-N-----N-----144		Qy	251	-----IW-----T-----K-D-G-N-----257	
Db	1140	DGSLTIPNVNEAHRGFTCHAQNAAGNDTRVTLTHTVHTPTINAEQEKIALQNDIVLE	1199	Db	2220	FKQIILWRNYOPIDLEAEDARITRLSNDRLTLILNVTENDEGQYSCRVKNDAGENSFDF	2279
Qy	145	-----HT-----D-----MMB-----V-----D-152		Qy	258	-----L-----A-----S-TL-----GO-H-----265	
Db	1200	CPAKALFPVRLWTEGEKIDSQI1PHTIREDGALVQNKLENTGVFVQCVSNLAGEDS	1259	Db	2280	KATVLPPTIIMLDKDKNKTAVEHSTVTLSCPATGKPEPDITWFKDGEAIHINIADIIP	2339
Qy	153	-----G-DV-----EIP-----P-----NK-AV-----163		Qy	266	-----K-G-P-----I-----269	
Db	1260	LSYTLVHEKPIIIEVPGVVDVVKGFTIEIPCRATGVPEVIRVTWNGIDLMKDEKFS	1319	Db	2400	NERVVISCPVYARPPAKITWLKAGKPLQSKFVKTSANGQKLYFLKLRDSSSKYTCIAT	2459
Qy	164	V-----LR-----G-----H-----E-----S-----170					

QY 270 -----F-----A-----L-----K-----W-----N----- 275
Db 2460 NEAGTDKRDPKVSMVLVAPSFDEPNVIRITVNSGNPSTLHCPAKGSPSPITITWLDKGNAI 2519
QY 276 -----K----- 276
Db 2520 EPNDRYVFFDAGRQLQTSKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPVIDGERREA 2579
QY 277 -----K-G-N-----FI-----LS----- 283
Db 2580 VAVIEGSELSFCDNSNSTGVVDVWQDGLTINQDTLRGDSFIQIPSSGKKMSFLSARKSD 2639
QY 284 -----AG-----V-----DK-T-----TI-----I----- 292
Db 2640 SGRYTCIVRNPAAGEARKLPDFAVNDPPSISDELSSANIQTIVPYVPVEINCVVSGSPHPK 2699
QY 293 --W-----D--A-----H-----T-----G----- 298
Db 2700 VYMLFDDKPLEPDSAAVELTNNGETLKIVRSQVEHAGTYTCEAQNNVGKARKDFLVRVTA 2759
QY 299 -----E-----AK-----Q----- 302
Db 2760 PPHFEKERBEVWARVGTMLLTCASSSVPLSSVYMHADHESVQNGVITSKYAANEKTLN 2819
QY 303 -----Q-----F----- 304
Db 2820 VTNIQLDDEGYFYCTAVNEAGITKPKFLIVETPYFLDOOKLYPIILGKRLTDCSATG 2879
QY 305 --P--F----- 306
Db 2880 TPEPTILFMKGKRLNESDEVDIIGSTLVIDNPQKEVGRYTCIAENKAGRSKDMWVEV 2939
QY 307 -----H-----S-- 308
Db 2940 LLPKLSKEWINVEVQAGDPLTLECPIDTSGVHITWSRQFGKGDMDRAQSSDKSL 2999
QY 309 --A-P--A--L-----D-----V-D----- 315
Db 3000 YIMQATPEDADSCTIANDAGAEAVQVTVNTPPKIFGDSPTSTEIVADTTLEIPCRT 3059
QY 316 -----W-----Q-- 317
Db 3060 EGIPPEISWFLDGKPILEMPGVYTKQDLSRIDNKPQEGRYTCVAENKAGRAEQDT 3119
QY 318 -----S-----N-----T-----F----- 322
Db 3120 YVEISEPPRVVMASEVVRVVEGRQTTIRCEVFGNPEPVNWLKDGPEYTSDDLQFSTKLS 3179
QY 323 -----A-----SC-----S-----TD-----M-C- 330
Db 3180 YLHLRETTLDGGTYTCIATNKAGESQTTDVEVLVPPRIEDERVLQKEGNTYVWHQ 3239
QY 331 -----I-----HV-----C-----K 335
Db 3240 VTGRPVYVTKRNGKEIEQFNPLHNRATRADEKYSCTASNEAGTAVADFLIDVFTK 3299
QY 336 -----L--QO-----D-----RP-----I-K- 343
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRPPNMDNIILSPRGDTLMILKA 3359
QY 344 -----T-----P-----Q----- 346
Db 3360 QRPDGLGYTCVATNSYGDSEQDFKNVYTKPYIDETIDQTPKAVAGSILKCPVLGNPT 3419
QY 347 -----G-----HT-----N-----E-----VN-----A--I- 355
Db 3420 PTVTKRGDDAVPNSRHTITVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479
QY 356 -----K-----W-----D-----P-----T- 360
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDGSHITI 3539

QY 361 -----G-----N-----L-----LA-----S-----C--S 368
Db 3540 NKAKLSDGKYYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYLECPI 3599
QY 369 -----D-----D-M----- 371
Db 3600 GIQPDVIVTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGGKASHDF 3659
QY 372 -----TL-----K-----I-----W-----S-----M-- 378
Db 3660 SLDLVSPPEFDIHGTOPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKDSRALDGLDLD 3719
QY 379 -----K-----Q-----D----- 381
Db 3720 NVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPVIDISRNDVQPV 3779
QY 382 --N-----C-V--H-----D-----L-----Q-----Q-----H----- 390
Db 3780 AVNQPTIMRCAVTGHPPFSIKMLKNGKEVTDDEINIRIVEQGVQLILRTDSDHAGKMSCV 3839
QY 391 --N-----KE-----I--Y--TI-----KW----- 399
Db 3840 AENDAGVKEMLVDVFTPPVSVKSDNPIKALGETITLFCNASGNPYPOLKWAKGSLI 3899
QY 400 --SP-----T--G-----P-----GT-----N----- 407
Db 3900 FDSPDGARISLKGARLDIPLHKKTDVGDYTCQALNAGTSEASVSDVLVPPPEINRDGID 3959
QY 408 -----N-----P----- 409
Db 3960 MSPRLPAQOSLTLQCLAQKVPQMRWTLNGALTHTSTPGITVASDSTFIQINNVSLSK 4019
QY 410 -----N-A--NLM-----LA-----S-- 417
Db 4020 GVYTCVAENVAGSDNLMYVVDVQAPVINSNGTKQVIEGELAVIECLVEGYPAPQVSWLR 4079
QY 418 -----A-SFDS----- 422
Db 4080 NGRVETGVQVRYVTDGRMLTIEARSLSDSGIYLCSATNEAGSAQAQAYTLEVLVSPKII 4139
QY 423 --T-----VR-L-----W-----DV----- 429
Db 4140 TSTPGVLTPSSGSKFSLPCAVRGYPDPPIISWTLNGNDIKDGENGTIGADGTLHIEKAE 4199
QY 430 -----D-----R-----G-----I-C-I----- 435
Db 4200 RHLIVECTAKNDAGADTLEFPVQTVIAPKISTSGNRYINGSEGTETVIKEIESESSEFS 4259
QY 436 -----H-----T-L-----T-----K 440
Db 4260 WSKNGVPLLPNNLIFSEDYKLIKILSTRLSDOGEYSCTAANKAGNATQKTNLNVGVAPK 4319
QY 441 -----H--Q-----E-----P-V--Y-----SVAF-- 450
Db 4320 IMERPRTQVHVKGQDQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTTKKKSVIFSSI 4379
QY 451 SP-----D-----G----- 454
Db 4380 SPSQAGVYTCBAENVVASTEEDIDLIVMTPPEVPMNVSTNPROTVFLSCNATGIPEP 4439
QY 455 -----R-----Y-----LA-----SGSF-----D-----K 464
Db 4440 VISWNRDSNIAIQNNEKYQILGTLTILAIRNVLPDDG--FYHCIAKSDAGQKIATKILVNK 4498
QY 465 -----C-----VHI-----W-----N----- 470
Db 4499 PSDRPAPIVWECEDEKPKPKTEYMDRGDTPDDNPOLLPKDVEDSSLSNGSIAYRCMPGP 4558
QY 471 -----T-----Q-----T--CA-----LV----- 477
Db 4559 RSRTVLLHAAPQFIVKPKNTAAIGAIVELRCSAAGPPHPTITWAKOKLIEDSKFEIA 4618
QY 478 --H-----S----- 479

Db 4619 YSHLKVTILNSTSDSEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAVITCYE 4678
Qy 480 ---Y-RG-T---G---GI--- 486
Db 4679 RNQAYSRLTWEYNGVMPKPLAGIHFMWNSGLVILDTSSLKEGDLLEYLTCKVRNRRHS 4738
Qy 487 ---FE--- 493
Db 4739 IPHLTSAFEGVEPKTIDKVEVNGSDSVLDCVTSPLTHVVTNKDQKMLDDALYV 4798
Qy 494 ---A---G---D---KV---G---AS--- 501
Db 4799 LPNNSLVLLNVEKYDEGVKCVASNSIGKAFDDQLNLYGGSSRREAYKKNEDASTTTI 4858
Qy 502 ---A--- 505
Db 4859 TTTSPTTTTTETPLTTIIPALITLPAKOYPTDDYHEGSANDDGFPTQDSLFEFNPPL 4918
Qy 506 ---SV---CV--- 511
Db 4919 HPEISVNTDCAGTINENGDCVDKDKTHNLKILTGHNHCPEGPMNPHTRICEDLDECA 4978
Qy 512 ---L--- 514
Db 4979 FYQPCDFECINYDGGFCNCPLGYELAEGRDWNCECVRCEDGK 5024

RESULT 2
T43290
hemictentin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T43290; T20993; T24734
R:Voegel, B.E.; Hedgecock, E.M.
A:Submitted to the EMBL Data Library, June 1998
A:Description: Hemictentin is required for hemidesmosome mediated cell adhesion and germ-
A:Reference number: 222396
A:Accession: T43290
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5198 <WOG>
A:Cross-references: UNIPROT:O76518; EMBL:AF074901; PIDN:AAC36792.1
R:Sulston, J.
A:Reference number: 219355
A:Submitted to the EMBL Data Library, December 1994
A:Accession: T20993
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <WIL>
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CBSP:F15G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
A:Submitted to the EMBL Data Library, December 1994
A:Reference number: 219929
A:Accession: T24734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CBSP:F15G9.4b
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: him-4; F15G9.4b
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
A:Exons: 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;
A:Indels: 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/1

Query Match 70.9%; Score 2595.3; DB 2; Length 5198;
Best Local Similarity 9.0%; Pred. No. 7.2e-53;
Matches 453; Conservative 50; Mismatches 9; Indels 4537; Gaps 389;

Qy 1 M-----SI-----SS-----D-----E----- 7

Db 1 MGRSPWLYGVGLGILLIATTCSSVNDKNDPTGKSSLAFAVFDITGSMFDDLVQREGAAK 60
Qy 8 ---V---N---F---L---VY--- 13
Db 61 IFKTVMAQREKLYNYIMVFFHDPYLGEIINTTDSYFMRQLSKVYVHGSGDCPEKTLTG 120
Qy 14 ---R---Y---L---QF---SG---F---SHSAF-T 27
Db 121 ILKALQISLPSSFIYFTDARSXDYHLEDEVLTIOEKOSVVVFMVMTGCGNRTHGFGFT 180
Qy 28 ---FG-I---KS---H-ISO---S-NI--- 39
Db 181 YEKIAAASFGQPHLEKSDSVTVLEVRHAYQKKVHLMYEAERGGTVSRNIPVDKHL 240
Qy 40 ---N---GA---LV---P-P--- 46
Db 241 ELTISLSGDKDDNDLDIVLRDPEGRVTKRLYSKEGGTIDLKNVKILRLKDPSPGVWTV 300
Qy 47 ---A---LI---S-I--- 52
Db 301 NNTSRLKHTIRVFGHGAVDKYGFAFRPLDRIELARPRVPLNQDTYLLINMTGLIPGTV 360
Qy 53 ---I---Q---KGL---Q-Y---VE---AE- 63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYDEDNYEFMIAPT 420
Qy 64 ---V---S-I---N---ED---GTLF--- 73
Db 421 AIGSVIVGGPRAFMSPIHQEFVGRDLNLSCTVESASAYTIYVVKTGEDIIGGLFVHN 480
Qy 74 ---D---G---R---P-I---E--- 79
Db 481 TSVWITPELSLKDAGEYECRVISNNGNYSVKTRVETRESPEIFGVNRNVSPVLEGAFLH 540
Qy 80 ---S---L---SL-I---DA--- 86
Db 541 CSTRSAGEVEIRWTRYGAIVFNGPNTERTNGTLKIHVTRADAGVYECMAKNAGMST 600
Qy 87 ---VM---P-DV---V- 92
Db 601 RKMRLDIMEPPSVKVTQDVIYFNMREGVNLSCAMGDKPKVHVYFKGRHLLNDYKQVG 660
Qy 93 Q---T---R---Q---QAYRD---KLA---O--- 105
Db 661 QDSKELYIRDATHDEGYECRAMSQAGQA-RDITDMLATPKVEIIONKMMVGRDRV 719
Qy 106 ---Q---Q---A---A---A---A- 111
Db 720 SPECKTIRGPHPKIRWPKNGKDLIKPDDYIKINEQLHMGAKDEADAGAYSCVGENMAG 779
Qy 112 ---A---A---A---A---A---A- 114
Db 780 KDQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATNPN 839
Qy 115 ---A---A---A---ASO--- 120
Db 840 RYQLADGNLLITDAQIEDQGOFTCIARTYQQSQSSTLMTGLVSPVLGHVPPEEQLI 899
Qy 121 ---Q---GS--- 123
Db 900 EGQDLTLCVVVLGTPKPSIVMIKDKPVEEGPTIKIEGGGSLRLRGGNPKDEGKYTCI 959
Qy 124 ---A---K---N---GE--- 128
Db 960 AVSPAGNSTLHINVLIIKKPEFVYKPEGIVFKPTISGMDEKXHAVVNSTHVDLGEQFA 1019
Qy 129 ---NT---A-N-G- 133
Db 1020 IPCVSGTTPPIITWVLDGRPTTPNSRDTFTVADNTLIVRKADKSYGVYTCQAINSGD 1079
Qy 134 -E-----N- 136

Db 1080 NEOKTIRIMTPIWISGOSFNMVDDLTFTIPCDVYGPDPKPVITWLLDDKPFTEGVVNE 1139
QY 137 -G-----AH-----TI-A-N-----N----- 144
Db 1140 DGSLLTIPVNEARGTFTCHAQNAAGNDRTVTLTHTPTTNAENQEKIALQNDIVLE 1199
QY 145 -----HT-----D-----MMB-----V-----D- 152
Db 1200 CPAKALPPVPVRLWYBEGEKIDSLPHITIREDCALVQNVKLENTGVFCQVSNLAGEDS 1259
QY 153 -----G-DV-----EIP-----P-----NK-AV----- 163
Db 1260 LSYLTWVHBKPIISEVPGVDVWVGKFTIEIPCRATGVBEVIRTNKNGIDLKMDKKFS 1319
QY 164 V-----LR-----G-----H-----E-----S----- 170
Db 1320 VDNLTGLRIEADKNDIGNVNVVNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGDR 1379
QY 171 -----E-----V-F-----G-----H-----E-----S-----I-C----- 175
Db 1380 VELKCYVEASPASVTWFRGIAIGTDTKGYVVESDGLTIVQASVEDATIYTCASNEA 1439
QY 176 -----A----- 176
Db 1440 GKAEANLQVTVIASPIDIKDPDVVTOBSIKESHFSLYCPVFNPLPQISWYLNKPLIDD 1499
QY 177 -----W----- 178
Db 1500 KTSWKTSDDKRLHVFKAKITDSGVYKCVARNAGSGSKFQVEVIVPLNLSKYYKKV 1559
QY 179 -----PVS----- 181
Db 1560 FAKEGEBVTLCPSVGPFPVQPNVWVDGTWBERKKYKGATLSNDGLTLHFDVSVKQEG 1619
QY 182 -----D-L-LA----- 186
Db 1620 NYHVAQSGKNILDDIVLSVLAVPIVGBDDNLEVLFGKDISLSCDLQTESDDKTFVWS 1679
QY 187 -----GS-----G-----DS-----T-----A-R----- 194
Db 1680 INGESDRPDNVQIPSDGHLRYITDAKPNNGKYMCRVNSAGAKARTLTLDVLEPPV 1739
QY 195 -----EN-----IW-----N-----LS----- 199
Db 1740 EPVFEANQKLGNNPIILQCVTGNPKPTVIWKIDGNDVDKSWLPDESLSLLRIEKLTKG 1799
QY 200 -----EN-----S-----T-----S-GS-----T- 207
Db 1800 SAQISCTAENKAGTASRDPFIQNIAPTAKNEGDOETIIPRESEITITLDCPVSLGDFQITW 1859
QY 208 --Q-----L-VL--R-H--C----- 1916
Db 1860 MKQGLPLTENDALFTLDNLTILNANRDHEDIYTCVANNTAGQVSKOPDVVVQVLPKIK 1919
QY 217 -----EG-----G-----Q-DVP----- 223
Db 1920 NAVVTLIINEGEBEILTCDAEGNPTAKWDFNQGLPKAEAVFVNNHTVNNVTKYHT 1979
QY 224 -----SNK-----DV-----T-S-L-----D----- 232
Db 1980 GYVKYATNKVGQAVKTIINHVRTKPRFESGLTESELTVNLTRISITLECDVDDAIGVGIS 2039
QY 233 W-N-----SE-G-TL-----L-----A-T-----GSY----- 245
Db 2040 WTUNGKPLAETDGVQTLAGRPLHIVSAKTDHDSIYACTVTNEAGVATKTNLFVQVPP 2099
QY 246 -----DG----- 248
Db 2100 TIVNEGGEYTVIENNSVLVPCVETGKPNPVWTKDGRPVGDLKSVQVLSGQOKFVHA 2159
QY 249 -----A-----R----- 250
Db 2160 EIAHKGSIYICMAKNDVGTAEISFDVDIIITRPMIQIKINIVTAIGKALPFKCPIDDDKN 2219

QY 251 -----IW-----T-----K-D-G-N----- 257
Db 2220 FKQIILWLNYPIDLEADARITLNSNDRRLTILNVTENDEGOYSCRVKNDAGENSFDF 2279
QY 258 -----L-----A-S-TL-----GO-H----- 265
Db 2280 KATVLVPPTIIMLDKDKNKTAVEHSTVTLSCPATGKPEPDITWFKDGEAIHINIADIIP 2339
QY 266 ----- 265
Db 2340 NGELNQNQLKITRIKEDGAKYTCADNSAGSVQDVNVNVTIPKIEKDGIPSDYESQ 2399
QY 266 -----K-G-P----- 269
Db 2400 NERVVISCVPYARPAKITWLGKAGPLQSDKFVKTSANGQLVLPKLRDSSKYTCIAT 2459
QY 270 -----F-----A-----L-----K-----W----- 274
Db 2460 NEAGTDKRDVKVSMVAFPSDFEPNIRRVITVNSGNPSTLHCPAKGSPSPITITWLKDGNAI 2519
QY 275 --N-----K----- 276
Db 2520 EPNDRYVPFDAGRQLQISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPIDGERREA 2579
QY 277 -----K-G-N-----FI-----LSA----- 284
Db 2580 VAVIEGFSSELCDSNSTGVVQKDLGTINQDILRGDSFIQIPSSGKMKFSLARKSD 2639
QY 285 -----G-----V-----DK-T-----TI-----I----- 292
Db 2640 SGRYTCIVRNPAAGEARKLPDFAVNDPPSISDELSSANIQTIVPYVPVEINCVVSGSPHPK 2699
QY 293 --W-----D-A-----H-----T-----G----- 298
Db 2700 VYMLFDDKPLEPDSAAYELTNNGETLKIVRQVEHAGTYTCEAQNNGVKARKDFLVRVTA 2759
QY 299 -----E-----AK-----Q----- 302
Db 2760 PPHPEKEREVVARVGDWMTLLTCNAESSVPLSSVYVWHAHDESVQNGVITSKYAANEKTLN 2819
QY 303 --Q-----F----- 304
Db 2820 VTNIQLDDGFFYCTAVNEAGITKFFKLVITPFLDQOKLYPIILGKRLTLDCSATG 2879
QY 305 --P--F----- 306
Db 2880 TPTPTILFMKQKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVEV 2939
QY 307 -----H----- 308
Db 2940 LLPPKLSKEWINVEVQAGDPLTLECPIEDTSGVHITWSRQFGKDGQOLDMRAQSSDKSL 2999
QY 309 -----A-P-A-L-----D-----V-D----- 315
Db 3000 YIMQATPEDADSYCIAVNDAGBAVFOVTNTPPKIPGDSFSTTEIVADTTLEIPCR 3059
QY 316 -----W-----Q----- 317
Db 3060 EGIPPEISWFLDGKPILEMPGVITYKQGLSURIDNIKPNQEGRYTCVAENKAGRAEQDT 3119
QY 318 -----S-----N-----N-----T-----F----- 322
Db 3120 YVEISEPPRVWASEVMRVVEGRTTIRCEVFGNPEPVVNLKDGEPYTSDDLQSTKLS 3179
QY 323 -----A-----SC-----S-TD-----M-C- 330
Db 3180 YLHLRETTLDGTVTCIATNKAGESQTTDVEVLVPPRIEDEERVLOGKEGTYVWHQ 3239
QY 331 -----I-----HV-----C-----K 335
Db 3240 VTGRPVVYVWKRNGKIEQFNPVLHIRNATRADEGKYSIASNEAGTAVADFLIDVFTK 3299

```
QY 336 -----L-CQ-----D-----RP-----I-K- 343
Db 3300 PTFETHETFNIVEGSAKIECKIDGHPKPTISWLKGRPFNMNDNIILSPRGDTLMILKA 3359
QY 344 -----T-----F-----Q----- 346
Db 3360 QRFDGGLYTCVATNSYGDSEQDFKVVVYKPIDETIDQTPKAVAGGEIILKCPVLGNPT 3419
QY 347 -----G-----HT-----N-E-----VN-----A-I- 355
Db 3420 PTVTWKRGDDAVPNDSRHTIVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479
QY 356 -----K-----N-----W-----D-----P-----T- 360
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFERGKPVVLYDRYSISPDGSHITI 3539
QY 361 -----G-----N-----L-----LA-----S-----C-S 368
Db 3540 NKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPIS 3599
QY 369 -----D-M----- 371
Db 3600 GIPOPDVITWTKNGMDINWTSRVILAQNNETFGIENVQVTDQGRYTCTATNRGKASHDF 3659
QY 372 -----TL-----K-----I-----W-----S-----M-- 378
Db 3660 SLDLVSPPEFDIHTGTOPTIKREGDTITLTCPIKLAEDIAQVMDVSVTKDSRALDGLDLD 3719
QY 379 -----K-----Q----- 381
Db 3720 NVDISDDGRKLTISQASLENAGLYCTIALNRAGEASLEFKVEILSPPIDISRNDVQPOV 3779
QY 382 -----C-V-H-----D-----L-----O-----H----- 390
Db 3780 AVNQPTIMRCVATGHPSPSIKWLKNGKEVTDENIRIVEGQVLQILRTSDHAGKWSCV 3839
QY 391 -----N-----KE-----I-Y-TI-----KW----- 399
Db 3840 AENDAGVKELEMDLVDFTPPVWSKSDNPIKALGETITLPCNASGNPYQLKWKAGGSLI 3899
QY 400 -----SP-----T-G-----P-----GT-----N----- 407
Db 3900 FDSFDGARIKLGARLDIPLHKKTVDGYTCQALNAAGTSEASVSVDVLVPPPEINRDGID 3959
QY 408 -----N-----P----- 409
Db 3960 MSPRLPAQOQLTLQCLAQKFPVQMRWTLNGTALTHTSTPGITVASDSTFIQINNVSLSDK 4019
QY 410 -----N-A-----NL-----LA-----S----- 417
Db 4020 GVTYCAENVAGSDNLMLNVDVVPVQAPVINSNGGTQKQVIEGELAVIECLVEGPAQVSWLR 4079
QY 418 -----A-SFDS----- 422
Db 4080 NGNRVETGVQVRYVTDGRMLTIEARSLDSGLYLCASANEAGSAQAQVTLVLVSPKII 4139
QY 423 -----T-----VR-L-----W-----DV----- 429
Db 4140 TSTPTGVLTPSGSKFSLPCAARGVPDPPIISWTLNGNDIKDGENGTIGADGTLHIEKAE 4199
QY 430 -----D-----R-----G-----I-C-I----- 435
Db 4200 RHLIYECTANDAGADTLFPVQTVIAPKISTSGNRYINGSEGTETVIKCEISESBSFS 4259
QY 436 -----H-----T-L----- 440
Db 4260 WSKNGVPLPSSNLIPESEYKLIKILSTRLSQGEYSCTAANKAGNATQTNLNVGVAPK 4319
QY 441 -----H-Q-----E-----P-V-Y-----SVAF--- 450
Db 4320 IMERPRTQVVKHQDQVTLWCEASGVQPAITWKDNELLTNGVDETATTKKKSIFSSI 4379
QY 451 SP-----D----- 454
```

```
Db 4380 SPSQAGYVTCYTKAENWVASTEEDIDLVIWIPPEVPPERMNVSTNPRQTVFLSCNATGIPEP 4439
QY 455 -----R-----Y-----LA-----SGSF-----D-----K 464
Db 4440 VISWEMDSNIAIONNEKYQILGTTILAIRNVLPDDDG-FPHCTIAKSDAGOKIATRKLIYNK 4498
QY 465 -----C-----VHI-----W-----N----- 470
Db 4499 PSDRPAPIWVECDCEKKGKPKTEYMDRGDTPODNPOLLPKWQVDESSLSNGSIAYRCMPGP 4558
QY 471 -----T-----Q-----T-----GALV----- 477
Db 4559 RSRTVLLHAAPQFIVKPKNTTAAGIAVELRCSAAGPPHPTITWAKOGKLIEDSKFEIA 4618
QY 478 -----H----- 479
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAVITCYE 4678
QY 480 -----Y-RG-T-----G-----GI----- 486
Db 4679 RNQAVSRGLTWBYNGVPMFKNLGIFHMNNGSLVILDTSSLKEGDELYTCKVRNRRHS 4738
QY 487 -----FE-----V-C-----W----- 491
Db 4739 IPHLSAFEQEVKTIIDKVEVNGDSVVDCEVTSDDLTHVWTKNDQKMLDDDAIYV 4798
QY 492 -----N-----A-----A-----GD-----K-----V----- 498
Db 4799 LPNNSLVLLNVBEKYDEGVYKCVASNSIGKAPDQTQNLNVEGDFLPTGFEGSGGINIDSS 4858
QY 499 -----G-----AS----- 502
Db 4859 NAGGSSREAYKKNEDASTTTTTSPTTTTTPPLTTIIPALITLPAKQYPTDDYHE 4918
QY 503 -----S-----DG-----SV-----CV----- 509
Db 4919 GSANDDGFGPTQDSLFBFNPPLHPEISVNTDCAGTINENGDCVDKDGKTHNLKILIGE 4978
QY 510 -----LD----- 512
Db 4979 NHCPEGFAMNPHTRICEDLDECAFYQPCDFECINVDGFGQCNCPLGYELABEGCRDNEC 5038
QY 513 -----R-----K 514
Db 5039 ESVRCEDCK 5047

RESULT 3
I38346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C:Accession: I38346
R:Labelit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38346
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:gl017426; PIDN:CAA62189.1; PID:gl017426
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 70.9%; Score 2594.7; DB 2; Length 7962;
Best Local Similarity 8.7%; Pred. No. 1.8e-52;
Matches 450; Conservative 52; Mismatches 9; Indels 4663; Gaps 382;
QY 1 MS-----IS-----SDE-----V-N----- 9
```

Db	57	MTQFEDTYOLEIAEAYPEDEGTYTFVANNVAVQVSTANLSLEAPESILHERIEQBIEME	116	Db	1136	VGSSICARSVTLREPPSPFIKKIESTSSLRGGTAAFOATLKGSLPITVTWLKDSDEITEDD	1195
Qy	10	-----F-----L-----VY-----R--Y--	15	Qy	112	-----A-----A-----A-----A-----A--	119
Db	117	MKAAPVIRKIBPLEVALGHAKFTCEIOSAPNRFOWFKAGREIYESDKCSIRSKYIS	176	Db	1196	NIRMTFENNVAISLYLSGIEVKHDKYVCOAKNDAGIQRCSALLSVKEPATITEAVSIDV	1255
Qy	16	-----L--Q-----E-----S-----G-----F--	21	Qy	120	-Q-----Q-----G-----SAK-----	125
Db	177	SLEILATQVDCGEYTKASNEGVSVCTATLTVEAYPTFLSRPKSLTTFVGKAAFI	236	Db	1256	TQGDPA TLQVKSGETITAKWFKDQBELTGLSKYKISVTDVTSLIILKISTEKKDSGEYT	1315
Qy	22	-----S-----H-----	24	Qy	126	-----N-----G-----	127
Db	237	CTVTGTFVIETIWOKDGAALSPSPNWRISDAENKHILELSNLTIOGRGVSCASNKFGA	296	Db	1316	FEVQNDVGRSSCKARINVLDLIIPSPFTKCLKMDSIKGSFIDLECIVAGSHPIQWPK	1375
Qy	25	-----A-----F-----T-----	27	Qy	128	-----E-----NTA-----N-G-----E--	134
Db	297	DICQABLIIIDKPHFIKELEPVQSAINKKVHLECOQVDEDRKVTVTWSKDGOKLPPGKDYK	356	Db	1376	DOEISASEKYKPSFHDNTAFLEISOLEGTSGTYTCSATNKAGHQCSCGHLTVKPPPYF	1435
Qy	28	-----F-----G-----IK-----	31	Qy	135	-E-----N-----G-A-----H-----	139
Db	357	ICFEDIATLEIPLAKDKSGTVCTASNEAGSSCSATVTVREPPSFVKVDPSYMLP	416	Db	1436	VEKPOSQDVNPNTRVQLKALVGSTAPMTIKWPKDNKELHSGAARSVMKDDTSTSLFLFAA	1495
Qy	32	-----S-----H-----ISO-----SN-----IN-----G-----	41	Qy	140	-----T-I-----A-----N-----	143
Db	417	GESARLHCKLKGSPVI-QVTFKNKSELSESNTVRMVFVNSEAILDITDKVEDSGSYSC	475	Db	1496	KATDSGTIYCOLSNDVGTATSKATLTVKEPPQFIKKPSPVLVRNGQSTTFBCQITGTPK	1555
Qy	42	-A-----LV-----P-----PA-----AL-----IS-----	51	Qy	144	-----N-----H-----T-----D-----	147
Db	476	EAVNDVSGSDSCSTEIVIKPSPPIKLEPADIVRGTNALLQCEVSGTGPFSEISWFKDKKO	535	Db	1556	IRVSWYLDONEITAQKHGISFIDGLATQIISGARVENSCTVCEARNDAGTASCIELK	1615
Qy	52	-----I-----	52	Qy	148	-----MM--EV--D-----G-----D--V--	155
Db	536	IRSSKYRLFSQSLVCLFISFNSADVGEYBVCVANEVKGCCMATHLLKEPPTFVKV	595	Db	1616	VKEPPTFIRELAKPVEVVKYSDVELECEVTGTPPFEVTLKNNREIRSSKKTTLTDRVSFV	1675
Qy	53	-----I-----Q-----KG-----L-----Q-----	58	Qy	156	-----E--I-----PP-----N-----K-----	161
Db	596	DOLIALGGQVTQLQAARGSEPISTVMKGOEVIREDGKIKMSFNGVAVLIPDVQISF	655	Db	1676	NLHITKCDPDTGEXQCIVSNEGGSCSTRVALKEPFIKKIENTTTLVKSSATFQST	1735
Qy	59	-----Y-----VE--AE-----	63	Qy	162	-A-----V--V--L--R--GH-----ES-----	170
Db	656	GKTYTCLAENEAGSOTSVGELIVKEPAKTIERABLIQVTAGDPATLEYTVAGTPELKPKW	715	Db	1736	VAGSPISITWLKDDQILDDEDDNVVISFVDSVATLQIRSDNGHSGRYTCQAKNESGVER	1795
Qy	64	-----V--S--I--N-----E--D-----G-----T--L-----	72	Qy	171	-----E-----E-----V-----P-----	173
Db	716	YKDRPLVASKKYRISFKNNVAQLKFYSALHDSGQYTFEISNEVGSSCETTFVTDLDRD	775	Db	1796	CYAFLLVQEPAQIVEKAKSVDTVEKDPMTLECVAGTPELKVKWLKDGQIIVPSRYPFMS	1855
Qy	73	-----F-----D-----G--R-----P-----I--E--SL	81	Qy	171	-----E-----E-----V-----P-----	173
Db	776	IAPFFTKPLRNVDSVNGTCRLDCKIAGSLPMRVSWFKDGKEIAASDRYIAFVEGTASL	835	Db	1856	FENNVPRIQSVMKQDSQYTFKVENDFGSSSCDAYLRVLDQNIIPPSFTKLTMKDKVL	1915
Qy	82	SLI-----DA-----V-MP--DV-----V-----Q	93	Qy	174	-I--C-----A--W-----	177
Db	836	EIIRVDMNDAGNFTCRATNSVGSKDSSGALI VOEPPSFVTKPGSKDVLPGSAVCLKSTFQ	895	Db	1916	GSSIHWCKVSGSLPISAQWFKDGKEISTSAKYRLVCHERSVSLEVNLELEDTANYTCK	1975
Qy	94	-----T--R-----	95	Qy	178	-N-----P--V-----S-----D--	182
Db	896	GSTPLTIRFKGNKELVSGSCVITKEALESLELYLVKTSDSGTYTCKVSNVAGVECS	955	Db	1976	VSNVAGDACGILTVKEPPSPFLVKGROQAIIPDSTVFEKAILKGTFFPKIKWFKDDVEL	2035
Qy	96	-----Q--A-----Y-----	99	Qy	183	-----LL--AS-----GS-----	188
Db	956	ANLFFVKEPATFVEKLEPSQLKKGDATQACKVTGTPPIKITWFANDREIKESSKHRMSF	1015	Db	2036	VSGPKCFIGLEGSTSFLNLYSDASKTQYTCVTNDVSGSDCTTLLVTEPPKFKVKLE	2095
Qy	100	-----R-----D-----K-----	102	Qy	189	-----GDST-----A--R--W--N--L-----S-----E--N--ST--	203
Db	1016	VESTAVLRITDVIEDSGEYMCBAQNEAGSDHCSIIIVKESPYFTFKFPIEVLEKYDV	1075	Db	2096	ASKIVAGDSRLECKIAGSPBIRVWFRNEHELPAISKYRMTFFIDSVAVIQNNLISTED	2155
Qy	103	-----LA-----Q-----Q-----Q-----	107	Qy	204	SG-----ST-----	210
Db	1076	MLLAEVAGTPPPEITWFKONTILRSRKYKTFIQDHLVSLQILKFVAADAGEYQCRVTNE	1135	Db	2156	SGDFICEAQNAGPAGTSCSTKVIKPPVSSFPPIVETLKNAEVLSLECELSGTPPEVVM	2215
Qy	108	-----A-----AA--A-----	111	Qy	211	-----LR-----HC-----	214
Db				Db	2216	YKDKRQLRSSKKYKIASKNFHTSIHLNVDTSDIGYHCKAQNEVGSDTCVCTVKLKEPP	2275

QY 215 -----IRE----- 217
Db 2276 RFVSKNSLTVVAGEPAELQASIEGAQPIFVQWLKEKEVIRESENIRITFVENVATLQF 2335
QY 218 -----GG----- 219
Db 2336 AKAEAPANAGKYICQIKNDGGMRENMATLWLEPAVIVEKAGPMITVVGECTICLECKVAGT 2395
QY 220 -----QD----- 221
Db 2396 PELSVIEWYKDGKLLTSSQKHKFSYINKISSRLILSVERODAGTYTFOVQNNVGKSSCTAV 2455
QY 222 -----V-PS-----N-----K----- 226
Db 2456 VDVSRAVPFSFTRRLKNTGGVLGASCILECKVAGSSPIVAMFHEKTKIVSGAKYQTTF 2515
QY 227 -D-V-T-----SLD----- 232
Db 2516 SDNVCTLQLNSLSDSSDMGNYTCVAANVAGSDECRVLTVOEPPSFVKEPEPELVLPCKNV 2575
QY 233 -----W-----N-----S-E----- 236
Db 2576 TPTSVIRGTPFPKYNWFRGARELVKGDRCNIYFEDTVAELEFNIDISQSGEYTCVVSN 2635
QY 237 -G-----T-L-----A-----TG-----S-Y-----DGF----- 248
Db 2636 AQOASCTTRLFVKEPAAFLKRLSDHSVEPGKSILLESTYGTLPISVTWKOGFNITTS 2695
QY 249 -----A-R----- 250
Db 2696 KCMIVTTEKTCILEILNSTKRDAGQYSCIEIENAGRDVCGALVSTLEPPYFVTELEPLEA 2755
QY 251 -----I-W-----TK-----D-G----- 256
Db 2756 AVGDSVSLQOVAGTPEITVSWYKGTDLKRLPTPEYRTFTNNVATLVFNKNVINDSGEY 2815
QY 257 -----N-L-----AS-----TL-----G-----Q----- 264
Db 2816 CKAENSIGTASSKTVFRIQEROLPPSPARQKDIEQTVGLPVLTCRLNGSAPIQWCYR 2875
QY 265 -----H-----K-----G----- 267
Db 2876 DGVLLRDHENLQTSFVDNVATLKILQTLDSHQYSCSASNPLGTASSSARLTAREPKKS 2935
QY 268 -----P-----I-----F-----A-LK-----W-----NK----- 276
Db 2936 PFFDIKPVSDIVAGESADFECHVTGAQPMRITWSKDNKEIRPGNGYITTCVGNTPHLRI 2995
QY 277 -----KG----- 278
Db 2996 LKVGKDGSGQYTCQATNDVKDMCSAQLSVKEPPKFVKKLEASKVAKQGESIOLECKISG 3055
QY 279 -----N-FI-----L-----SA-----GV-D----- 287
Db 3056 SPEIKVSWPRNDSSELHESWKYNNFSFINSVALLTINEASAEADSGDYICEAHNGVGDASCST 3115
QY 288 -----K-----T-----T-----IIW----- 293
Db 3116 ALTVKAPPVFTQKPSVPVGLKGSVDILQCEISGTPPPFVVVWVKDKQVRNSKKFKITSKH 3175
QY 294 -D-----A-----H----- 299
Db 3176 FDTNLHLNLEASDVGEYHCKATNEVGSOTCSVKEKEPPRFVKKLSDTSLIGDAVEL 3235
QY 300 -A-----K-----Q----- 303
Db 3236 RAIVEGFQPIVSVWLKDRGEVIRESENTRISFIDNIATLQGSPEASNSGYICQIKNDA 3295
QY 304 -----F-----P-----F----- 306
Db 3296 GMRCSAVLTLEPARIIKEPBMVTTGNPFALCECVTGTPELSAKWFKDGRLELSADSK 3355

QY 307 -H-S-----A-----P-A-----L-----D-V----- 315
Db 3356 HHITFINKVASLKIPICAEMSDKGLYSFEVKNVSGKNCNTVSVHVSDRIIVPPSFIRKLKDV 3415
QY 316 -----W-----QS-----N----- 319
Db 3416 NAILGASVVLCECRVSGSAPISVCGWFOGNEIVSGPKQCSQSPSENVCTNLNLSLEPSDTGI 3475
QY 320 -----N-----TF----- 322
Db 3476 YTCVAANVAGSDRCSAVLTVOEPPSPQTPDSVEVLPGMSLTFTSVIRGTPPPFKVWFKG 3535
QY 323 -----ASC----- 325
Db 3536 SRELVPGESCNISLEDFVTELELFEVQPLESGDYSCLVNDAGSACTTHLFFVKEPATFV 3595
QY 326 -----S-----T-----D-----M----- 329
Db 3596 KRLADSFVETGSPVLEATYTGTPPISVSWIKDEYLIQSERCISITWTEKSTILEILEST 3655
QY 330 -----C-----I-----HV-----CKL-G----- 337
Db 3656 IEDYAOYSLIENEAGQDICEALVSVLEPPYFIEPLEHVEAVIGEIPATLQCKVDGTPAIR 3715
QY 338 -----Q-----D----- 339
Db 3716 ISWYKEHTKLRSAIPAYKQPKNVASLVINKVDHSDVGEYSCKADNSGVAVASSAVLVIK 3775
QY 340 -R-----P-----I-KT-----F-----Q-----G-----H----- 348
Db 3776 ARKLPFPFARKLVKHETLGFVPVAFECRINGSEPLQVSWYKDGVLKDDANLQTSFVHN 3835
QY 349 -----T-----N-----EV----- 352
Db 3836 ATLQILQTDQSHIGQYNCASNPLGTASSAKLILSEHEVPPFPDLKPVSDIALGESGT 3895
QY 353 -----NA-IK-----W-----D-----PTGN-----L-----L-----A-----SC-----S----- 368
Db 3896 FKCHVTGTAPIKITWAKONREIRPGNGYKMTLVENTATLVLKVGKGDAGQYTCYASNTA 3955
QY 369 -D-----DM-T-L-----KI-----W-----SMK-Q----- 380
Db 3956 GKDSCSAQLGVQVQPPRFIKKLEPSRIVKQDEFTREYCKIGSGPEIKVLWKBET-BIQES 4014
QY 381 -----D-----N-----C-----V-----H----- 385
Db 4015 SKFRMSFVDSVAVLEMHNLISVEDSGDYTCEAHNAAGSASSSTSLKVKEPPIFRKKPHPIE 4074
QY 386 -----D-----LQ-----Q-----H----- 390
Db 4075 TLKGADVHLECELOQTTPPHVSWYKDKRELRSKKYKIMSENFILSIHLNVDAADIGEY 4134
QY 391 -----N-----KEI-Y-----TI-----K 398
Db 4135 QCKATNDVGSDDTCVGSIALKAPPRFVKKLSDISTVVGKEVQLQTTIEGAEPISVVWFKDK 4194
QY 399 -----W-----S----- 400
Db 4195 GBIVRESDNIIWISYSENATLQFSRVEPANAGKYTCQIKNDAGMOECFATLSVLEPATIV 4254
QY 401 -P-----TG-----P-----G-----T-----N-----N----- 408
Db 4255 EKPEISKVTTGTCTLECTVAGTPELSTKWFKDGKELTSDNKYKISFFNKVSGLKLIINVA 4314
QY 409 P-----N-----A-----N-L-----M-----L----- 415
Db 4315 PSDSGVYSEVQNPVQKDSCTASLQVSDRTVPSPFTRKLKETNGLSGSSVWMECKYVGP 4374
QY 416 -ASAS-F-----DS----- 422
Db 4375 PISVWFHEGNEISSGRKYQTTLTONTCALTVNMLEESDSDGYTCIATNMGSDCSAPL 4434
QY 423 TVR-----L-----W----- 427

```
Db 4435 TVREPPSFVQKDPMDVLGTGNTVFTSIYKGTPTFPFSVMSKGSSELVPGDCRNVSLDSV 4494
QY 428 -----DVD--R-G--ICI-----H----- 436
Db 4495 AELELFDVTSQSGEYTCIVSNEAGKACTHLYIKAPAKFVKRLNDYSIEKGKPLILSG 4554
QY 437 -----T----- 437
Db 4555 TFGTPTPISVTKKNGINVTSPQRCNITTEKSPILEIPSSIVEDAGQYCNVIENASGKD 4614
QY 438 -----L-----TK----- 440
Db 4615 SCSAQILILEPPYFVKOLEPVPKVGSDSASLOCLAGTPEIGVSWYKGTDLRPTTTYKM 4674
QY 441 H-----OB-----P----- 444
Db 4675 HFRNNVATLVFNQVDINDSGEYICKAENSGEVSASTFLTVGEQKLPSPFSQRLRDVQST 4734
QY 445 V-YSVAF-----S-P-----DG-----R-----Y-- 456
Db 4735 VGLPVVFDCAISGSEPISVSWYKDGKPKLSDPNVQTSFLDNTATLNIKFTDRSLAGQYSC 4794
QY 457 LAS--GS-----PD-----KCH-----I--W-- 469
Db 4795 TATNPIGSSASSARLILTEGKNPPFPDIRLAPVDVAVGESADPEC-HVTGTQPIKVSNAK 4853
QY 470 -----N--T-----Q-T-----G-----A-----LVH 478
Db 4854 DSREIRSGGKYQISYLENSAHLTVLKVKGDSGQYTCYAVNEVGKDSCTAQLNIKERLIP 4913
QY 479 -----S-----Y----- 480
Db 4914 PSFTKRLSETVEETEGNFKLEGRVAGSQPIITVAVYKNNIEIQPTSNCEITKNNVLVLQ 4973
QY 481 R-G--TG-----G-----I-----F-----E-V--C- 490
Db 4974 VRKAGMNDAGLYTCVNSDAGSALCTSSIVIREPKPPVFDQHLTPVTVSEGEYVQLSCH 5033
QY 491 -----W----- 495
Db 5034 VQSGEPIQLWKAGREIKPSRCSFSPASGTAVLELRDVAKADSGDYVCKASNVAGSDT 5093
QY 496 -----DK-----V-G-----AS-----D----- 504
Db 5094 TKSQVTIKDKPAVAPATKKAANDGRLLFFVSEPSQIRVEKTTATFLAKVGSDPIPNKWT 5153
QY 505 -----GSV-----CV-----L----- 510
Db 5154 KGKWRQLNQGGRVFIHQKGEAKLEIRDTTKTDSGLYRCVAFNEHGEIESNVNLQVDERK 5213
QY 511 -----DLR---K 514
Db 5214 KOEKIEGDLRALMLK 5227
```

RESULT 4

I38344

titin, cardiac muscle [validated] - human

N:Alternate names: connectin

N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)

C:Species: Homo sapiens (man)

C>Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004

C:Accession: I38344; I38345; S20898; S20897; S20899; S63665; S73793

R:Labeit, S.; Gautel, B.

Science 270, 293-296, 1995

A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.

A:Reference number: A57430; MUID:96026330; PMID:7569978

A:Accession: I38344

A>Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL

A:Molecule type: mRNA

A:Residues: 1-26926 <LAB1>

A:Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:g1017424; PID:g1017425

R:Musco, G.; Tziatios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix m
A:Reference number: I38345; MUID:95119041; PMID:7819249
A:Accession: I38345
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:G602579; PIDN:CAA58243.1; PID:G602580
A>Note: conformation and properties are reported for a synthetic peptide corresponding t
R:Labeit, S.; Gautel, B.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380; PMID:1582406
A:Accession: S20898
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 13597-14200, 'I', 14202-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:G37192; PIDN:CAA45939.1; PID:G37193
A:Accession: S20897
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:G37190; PIDN:CAA45940.1; PID:G37191
A:Accession: S20899
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-22484
A:Cross-references: EMBL:X64697; NID:G37190; PIDN:CAA45938.1; PID:G37195
R:Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.
J. Mol. Biol. 256, 556-563, 1996
A:Title: Genomic organization of M line titin and its tissue-specific expression in two c
A:Reference number: S63665; MUID:96177761; PMID:8604138
A:Accession: S63665
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 26729-26825 <XOL>
A:Cross-references: EMBL:X92412; NID:G1236761
R:Gautel, M.; Leonard, K.; Labeit, S.
EMBO J. 12, 3827-3834, 1993
A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiat
A:Reference number: S37393; MUID:94008990; PMID:8404852
A:Accession: S37393
A:Molecule type: mRNA
A:Residues: 26831-26926 <GAU>
R:Improta, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A66736; PDB:1TIT
A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R:Pfuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A66201; PDB:1NCT
A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q32
C:Function:
A:Description: structural protein forming filaments in striated muscle
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; prot
C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein
F:24752-25008/Domain: protein kinase homology <KIN>
F:84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4068,4
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
F:16780,16976,17579,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18
21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,2484
F:26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental
Query Match 70.7%; Score 2585.4; DB 1; Length 26926;
Best Local Similarity 8.9%; Pred. No. 3.8e-51;
Matches: 435; Conservative 63; Mismatches 10; Indels 4366; Gaps 375;

QY 1 M-----SISSD-----EV--N-----F-----L-----11
Db 20024 MRNTEITAENLTINUKESVTADAGAYEYETAANSSGTTKAFINIVVLDPRPGPTGPPVIS 20083
QY 12 -----Y-----R--Y--L--Q--E-----S-----G-----20
Db 20084 DITEESVTLKWEPPKPYDGGSQVNYILLKRETSATVAVTEVSATVARTMMKWLTTGEEY 20143
QY 21 -----F-----S-----H-----SA 25
Db 20144 OFRIKAENRFGISDHIDSACVTVKLPYTPPGPSTPWNTVNTRESITVGMHEBPVNSGGSA 20203
QY 26 -----F--T-----F-----G--K--SH-----33
Db 20204 VUGYHLEMDKDRNSILWOKANKLVIRTHFKVTTISAGLIYEFVRYAENAAAGVKPSHPSE 20263
QY 34 -----N-----I--N-----G-----41
Db 20264 PVLAIACEPPRNVRITDISKNSVLSWQAPAFDGGSKITGYIVERRDDLPDGRWTKASFT 20323
QY 34 -----IS--Q--S-----N-----I--N-----G-----41
Db 20324 NVTETOFTISGLTQNSQYEFVARNVAGSISNPSEWVGPIITCIDSYGGFVIDLPLEYTE 20383
QY 42 -----A-----L--V--P--P-----A-----A--LI-----50
Db 20384 VVKYRAGTSVKLRAGISGKPAFTIENYKDDKELOTNALVCVENTTDLASILKADRLNS 20443
QY 51 -----S--I--Q--K-----G-----L--Q-----58
Db 20444 GCVELKLRNANASATIRVQILDKPGPGPIETKVTVAEKITLLWRPPADDDGGAKITH 20503
QY 59 Y-VEA-E-----VS-----I--NE-----D-----69
Db 20504 YIVERKTSRVVMSVSEHLEBECIITTKIKGENYIFRVRVANKYIGIGEPLESDSVAK 20563
QY 70 -----G-----T-----I-----L--F-----73
Db 20564 NAFVTPGPGIPEVTKITKNSMTVWSRPIADGGSDISGYFLEKRDKKSGLWFKVLKETI 20623
QY 74 D-----G-----R-----PI-E-S-LSL-----ID-----A-----86
Db 20624 RDTROKVTGLTENSQYQYVCAVNAAGQGFSEPF-YKAADPIDPPGPPAKIRIADST 20682
QY 87 -----V-----MP--D-----VW--Q-----T-----94
Db 20683 KGSITLGSKPVYDGGSAVTGYVVEIROGEEBETWTVTKGEVTRTEYVYVSNLKPQVNY 20742
QY 95 -R-----Q-----QA-----ID-----A-----98
Db 20743 FRVSANVCAGQGEPIEMNEPQAKDILEAPEIDLDVALRTSVAKAGEDVQVLIIPKGRP 20802
QY 99 -----YR-D-K-L--A-----Q-----105
Db 20803 PPTVTRKDEKNLGSARYSIENDSSLLTIPQVTRNDTGKYLFTIENGVEGPKSSTVS 20862
QY 106 -----Q--QA-----AA-----110
Db 20863 VKVLDTPAACQKLQVXKHSRGVTLLWDPPLIDGSPFIINYVIERKDATKRTWSVVSHKC 20922
QY 111 -----A-----AA-----A-----114
Db 20923 SSTSFKLIDLSEKTPFFFRVLAENEIGICEPCETTEPVAEVPAPIRDLMSKDKSTKTSV 20982
QY 115 -----AA--A-----AQ--Q-----121
Db 20983 ILSWTKPFDGGSVITEYVVERKGEQTSWAGISKTCEIYVSQLKEQSVLEFRVFAKN 21042
QY 122 --G-S-----AK-----N-----126
Db 21043 EKGLSDPVTIGPITVAKELIITPEVDLSDIPGAQVTVRIGHNVHLELPYKGPXKPSIWUK 21102

QY 127 -G-----EN--T-----A-----N-----132
Db 21103 DGLPLKESEFVRFSTENKITLSIRNAKKEHGGKYTVILDNAVCRIVAPITVITLGPSPK 21162
QY 133 --G-----E-----E--NG-----AHT-----IA-----142
Db 21163 PKGPFRFDEIKADSVILSWDVPEDNGGGEITCYSIEKRETSQTNMKWCVSSVARTTFKVP 21222
QY 143 N-----N-----H-----T-D-M-----148
Db 21223 NLVKDAEQFRVRAENRYGVSQLVSSIIIVAKHQFRIPOGPKPVINYNTSDGMSLTWDA 21282
QY 149 -----ME-----150
Db 21283 PVDGSGSEVTGFHVEKKERNLSILWQKNTSPISGREYRATGLVEGLDYOFRVYAENSAGL 21342
QY 151 -----VD--G-----DV--E--I-----PP-----NK--161
Db 21343 SSPSPSKFTLAVSPVDPGTPDYIDVTRETTILKWNPPRLDRGSGKIVGYSIEKROGNER 21402
QY 162 -----AV-----V--164
Db 21403 WVRCNFTDVSECOYTVTGLSPGRYEFRIIARNAVGTISPPSQSSGIINTRDENVPIVE 21462
QY 165 -----LR-----G-----167
Db 21463 FGPEYFDGLIIKSGESLRIKALVQGRPVPRVTWPKDGVIEKRMNWEITNVLSGTSLSFVR 21522
QY 168 -----H-----E-----S--E-----V--F--I-----174
Db 21523 DATDRHGVYTVAKNAGSAGAEIKVKVQDTPGKVGVPFRFTNITGERKMTLWMDAPLND 21582
QY 175 -C-----AW-----N-----178
Db 21583 GCAPITHYIIEKRETSRLAWALIEDKCEAQSVTAIKLINGEYQFRVSANVKGVRPLD 21642
QY 179 --PV-----SD-----L-----183
Db 21643 SDPVVAQIQTVPDAPGPEPSNITGNSITLTWARPESDGGSEIQOYILERREKSKTRWV 21702
QY 184 -----L-----ASG--S-----G-----189
Db 21703 KVISKRPISETRFKVTGLTEGNEYEFHVAENAGVPASGISRLIKCREPVNPGPPTV 21762
QY 190 -----D-S--T-----AR-----I-----W--N-----LS-----199
Db 21763 VKVTDTSKTTVLSKSPVFDGGMELIIGYIEMCKTDLGWHKVAEACVKTRYVTDLQ 21822
QY 200 --E-----N-----S--TSGS-----T-----Q-----L--209
Db 21823 AGEYKFRVSAINGAGKGDSCVET--GTIKAVDRLTAPELDIDANFKQTHVVRAGASIRLF 21881
QY 210 -----V-----LR--H-----C-----214
Db 21882 IAYQGRPTPTAVMSKPDNSLRLADITDTSFSTLTVCNCRNDAGKYTLTVENNSGSKS 21941
QY 215 -----I-----R-----EGG-----219
Db 21942 ITFTVKVLDTPGPGPIFTFKDVTGRSATLMDAPLDDGGARLHHYVVEKREASRSQVVI 22001
QY 220 -----Q-----D-----VP-----S--NK-----DV--TS 230
Db 22002 SEKCTROQIFKNDLAEGVPPYFRVSAVNEYGVGEVPEMPEIVATEQAPPRRLDVEDTS 22061
QY 231 -----L-----D-W-----233
Db 22062 KSSAVLWLPDHDGGSRTGYLLEMRQKSGDLWVEAGHTKOLTFTFVERLVEKTEYEFVR 22121
QY 234 --N-----SE-----G-T--LL--A-----T-----G-----243
Db 22122 KAKNDAGISEPREAFSSVIIKEPQIEPTADLTGINTQLITCKAGSPFTIDVPSIGRPAPK 22181
QY 244 -SY-----D-----G--F-----A-----249

QY 461 SF-----D-----K-----C-V-H- 467
 Db 24399 EFRVCENLGESEWSEIPEITPKSDVPIQAPHFKBELRLNRYOSNATLVCKVTGHP 24458
 QY 468 -I-W-----NTQ-----TG- 474
 Db 24459 KPIVKYRQKEIIADGLKYRIQRFKGGYHQLIIASVTDATVYQVRATN-QGGSVSGT 24517
 QY 475 -----A-----L-----VHSYRG-----T-----G----- 485
 Db 24518 ASLEVEVEPAKIHLPKTLKLEGMGAHVALRGVSVIKIPFSGKDPVITWQKQODLIDNNHY 24577
 QY 486 -I-----F-----E-----VC----- 490
 Db 24578 QVIVTRSTSLVFPNGVERKADAGFYVCAKMRFGIDOKTVELVDVADPPRGGVKSVDAS 24637
 QY 491 -----W-----N-----A----- 493
 Db 24638 RDSVNLWTPEPASPDSGSKIITNVIIVEKCATTAERMLRVQARETRYTVINLFGKTSYQFRV 24697
 QY 494 -A-----G-----DK-----V----- 498
 Db 24698 IAKNFKGSKPSEPEPTITKEDTRAMNYDEEVDETREVSMTKASHSTKELYEKYMA 24757
 QY 499 -----G-----AS-----A-----SD----- 504
 Db 24758 EDLGRGEFIVHRCVETSSKTYNAKFKVKGXTQVLVKKEISILNARHNILHLHESP 24817
 QY 505 -----G-----S-----VC-----V-LDLR- 513
 Db 24818 ESMEELVMIFISGLDIFERINTSAFELNREIVSVYHVQCEALQFLHSHNIGHDIRP 24877
 QY 514 -----K 514
 Db 24878 ENIIYQTRRSSTIK 24891

RESULT 5
 T30192
 Probable peptide synthetase - Aureobasidium pullulans
 C:Species: Aureobasidium pullulans
 C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
 C:Accession: T30192
 R:Peery, R.B.; Thornewell, S.J.; Tobin, M.B.; Skatrud, P.L.
 submitted to the EMBL Data Library, January 1997
 A:Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aureoba
 A:Reference number: Z20767
 A:Accession: T30192
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-10797 <PEE>
 A:Cross-references: UNIPROT:O94116; EMBL:U85909; NID:g4099310; PID:g4099313; PIDN:AAD005
 C:Genetics:
 A:Introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2
 C:Keywords: carrier protein
 F:1618-1688/Domain: acyl carrier protein homology <ACP1>
 F:3682-3752/Domain: acyl carrier protein homology <ACP2>
 F:5615-5685/Domain: acyl carrier protein homology <ACP3>
 F:7503-7573/Domain: acyl carrier protein homology <ACP4>
 F:7983-7972/Domain: acyl carrier protein homology <ACP5>

Query Match 70.6%; Score 2583.7; DB 2; Length 10797;
 Best Local Similarity 8.4%; Pred. No. 6.5e-52;
 Matches 459; Conservative 41; Mismatches 9; Indels 4923; Gaps 385;

QY 1 M-----SI-S-----S-----D----- 6
 Db 4302 MLNTQVIOQFGSGIGKSGILYMGVPTAAIHCTLPQFGVDLPAGTIGIPLDVTSCFIV 4361
 QY 7 -----E-V-----N-----FLV-----YR----- 14
 Db 4362 KPTESTKHASQLILPGEIGELVIGGHQLADGYLNREBQTRAAF-VTHPKFGGLYRTGD 4420

QY 15 -----Y-----L-Q-----E-----SG-----F-- 21
 Db 4421 KARLHRNTLECYGRISSSGQVQLRGQVELGEIEHAASKAGGCHAVIASVISGLLVLCI 4480
 QY 22 --SH-----SA-----F-----T-F----- 28
 Db 4481 GDHRVSSKDIKACQKNLPAWMPDSIVLLDDFFPLPSGVKDKKRLTDTYNSNTAQHVS 4540
 QY 29 G-----IKS-----H-----I-----SQ----- 36
 Db 4541 GSSDLSENAREITRIIESVLGVSIDHSTDLASAAGLSLRAIOVASQLRRQGCADLGALEL 4600
 QY 37 -----SNIN-----G----- 41
 Db 4601 LSVSNVLADELVRKADESININDNSEKWKQTVHELRSVERDFESKAFVSGIEDVLP 4660
 QY 42 -----A-LV-----P-P-----A-----AL-----IS- 51
 Db 4661 TPLQDAMLVETAKRPOAYCNELELTVPKIPVERVQALFALQORHTALRSRGMPSGVSH 4720
 QY 52 -----II-----Q-K-----G-----L-QY-----VEAE--V----- 64
 Db 4721 CAYTOVWKLTVTSQFAHVKSFTTGWSTNRETLLRPLHFQYKCSGAEILLVAIHALLY 4780
 QY 65 --S--I-----NE-----D--G- 70
 Db 4781 DOWSVEVILEDLTLNERTPERPSGAVNKFNLRRSEDTSHLDFWGEYLSVDTVGR 4840
 QY 71 -----TL-----F-----D- 74
 Db 4841 LPNLSPKMPPQPLQSIQHTIENDMETLQAASYSCHSAHFQAAAYAILLGYMGTEDT 4900
 QY 75 -----GR-P--IES-----LS-L--I-----DA-----VM----- 88
 Db 4901 VFCTVFSGRTPIVEIETESVMGPLLSTLPTPRINTLESRKFSVLSRLQEDNRKMRHSMTS 4960
 QY 89 -----P-----D-V-V-O-T-----R-----Q- 97
 Db 4961 LADIKACGFNPGAEVDFSIFVWQETARPDARAQTLNLNVEAHYLEFNLTELEPTQOG 5020
 QY 98 -----A-Y-----R-----DKL-----A--Q- 105
 Db 5021 VKTKATYQSSLLPLQHVKTLLQQLDALVKIIVARPETHNEISDQLPISVLSVANSEPOS 5080
 QY 106 -----Q-----Q-----Q- 107
 Db 5081 FVYKAGLSLVENHALNNSGGLALVFAHDIREGTSRMESLTYGELNTRANQLANLISQG 5140
 QY 108 A-----AA-----AA-----A----- 113
 Db 5141 AKRDELICVCMKESVSLYLILAAVKAGCGYLPVPETPAARIQILAEADVFKCLTDS 5200
 QY 114 -A--A-----A-----A----- 118
 Db 5201 MAPVIADVSRCHIMNVDTTDCSAQSGTGPQLDFKPTDIAVAVFTSTGTPKGVLTQEN 5260
 QY 119 -----S--Q--Q-----G-----SA-----K- 125
 Db 5261 ILSNLEVLKSIYPVPEGSLLQACNQAFDVSVEIFFTWYTGMLCSASKDVMFRDPEKA 5320
 QY 126 -N-----GE----- 128
 Db 5321 INELETHLSLTPVNAULTDPAHPRVKFLVTAGEAVTHVHVGAWAGKLYOYGPSETT 5380
 QY 129 N--T--A-----N-GE--EN-----GA----- 138
 Db 5381 NICTVNSAVESDHVINNIGPAPENTSFAVLTOGDDFQLVPLGGLGELCGGQQVFRGYQN 5440
 QY 139 -----H-----TI----- 141
 Db 5441 MPELTSKIINPNRYRSGDLGRLLPDGTILIQGRTDDQKRTGRIEUGEISGCLL 5500

QY 142 ---AN---N---H--- 145
DB 5501 QPFSVQNCALIEVKTADKERLMAFIPSGYSKDSYILOPKNLEIISIAHLADNLP 5560
QY 146 ---T---DM--- 148
DB 5561 AYMPDALVPVSAIPQTSQKIDKRRLASGSAITVEDNLNAYSRGADDDTSLSATEQQ 5620
QY 149 ---M-E--V---D---G---DV-EI--P--- 158
DB 5621 LASALADTLQMSQTSIGRSTSPFALGLDSVAIRLATNLRLKEYGIDVSOILKRTIAR 5680
QY 159 ---P---NK---AV---VL--- 165
DB 5681 LAPLLGGESSKOTNEPVTADCEAAVGYLHDSVWSQLHEHGTQVSOVLPCTPLQEAMLS 5740
QY 166 -R--G---H---E---S-E--V-PI--- 174
DB 5741 ARDTGSSAYRNKTLFSLHGSVDKLCACWEVMLQRHDIILRTIFLSTEDSRFPFVQAVLSQ 5800
QY 175 ---C---A---W--- 177
DB 5801 WTLPWQECDDIPDLSTLLDSAKAGGDSIVDHSPPMKIQVYRSESTVYLLLDHMHALYDA 5860
QY 178 ---N---PVS--- 182
DB 5861 NAMSNNLYEVEQLYKQDLSAPVSPKPLNFMISTSVBEADALFRDQLREFVPKPKRTD 5920
QY 183 ---L---LA--- 187
DB 5921 VKSGFTITGRNLNYSKPMVETFLSKHSTTMLSLTQAMWMTLAASQSYSDVCCGNVSGR 5980
QY 188 --- 187
DB 5981 SVPVDGIESLVAPCFNTIPVRVDLSKHSNGLVVKALQRWNIDSLPYQLTPLRRIQAAG 6040
QY 188 ---SG--D---ST--- 192
DB 6041 TNGKRLFDLSVLLQDDTDLDSAINRLEGESGMDMYTSGTGLPKAVCLSHRAVTSQSL 6100
QY 193 A---R---I--W---N--- 197
DB 6101 AHDRFIPSPSRFLQFASPTFDVSVFIFPPWYRGATLVSVERNRLGLDLPGTITSLNIDA 6160
QY 198 ---L---S---EN---S---T---SG--- 205
DB 6161 ABELTSVAASLVHRHENVTLRALLTIGEMLNTQVIOQFGSGSGILYGMVGTPEAAI 6220
QY 206 ---ST--- 209
DB 6221 HCTLPQFGVGLPAGTIGIPDVTSCFIVKPTESTKHAQLEILPIGEIGELVIGGHOLA 6280
QY 210 ---V---L---R-H---C--I---R--- 216
DB 6281 DGYLANREOTRAAFVTHPKFGLGYRTGDKARLHRNGTLECYGRISSGQVKLRQORVELGE 6340
QY 217 -E--GG---O---D-V--- 222
DB 6341 IEHASKAGGCHAVIASVISGLLVLCIGDPHRVSSKDIKSAQKWPAYMIPSDIULLD 6400
QY 223 ---P---SN---KD---V---T--- 229
DB 6401 DFPYLPSPGVKKRLETDNSNTAQHVSGSDLSENAREITRIESVLGVSIDHSTDLSA 6460
QY 230 ---SL---D---W--- 233
DB 6461 AGLDSLRAIQVASQLRROGCADLGALELLSVSNVLALDELVRKADESINNDSEKWKQ 6520
QY 234 ---N---S--- 235
DB 6521 TVHELRASSVERDFESKAFVSGIEDVLPCTPLQDAMLVETAKRPQAVCNELRLTVSPKIPV 6580
QY 236 E---G---TL---LA---TG--- 243

DB 6581 ERVQALFALQRTALRSGFMPGSHVCAYTQVIMKTLVTSQFAHVKSFTTGWSVTNRE 6640
QY 244 ---S---YD---G--- 247
DB 6641 TLLRPLHFQYKCSGABAILVAIHHAHYDQMSVEVILEDETLLQNERTPERSFGAVNK 6700
QY 248 -FA--R---I-W---TKD--G--NL--- 258
DB 6701 FPNLRSEDOTSHLDPWGEYLS-DVTPGRLPNLSPKWMPPOPQLOSTQHTIEMDMETLROA 6759
QY 259 ---A---S--TL---G--- 263
DB 6760 AHSYCSAHVFFQAAVAILLGFYMGTEDTVFGTVFSGRITPIVEIBSMVGPLLSTLPTRI 6819
QY 264 ---Q---H---K-G--P---IFALKW--- 274
DB 6820 NTLERKFSVDVLSRLQEDNRKIMRHSMTSLADIKKACGNPGEAVPDSIFV--WQETARP 6877
QY 275 ---N---K--- 276
DB 6878 DARAQTLLNLVEAHDYLEFNLTELEPTQOGVKTKATYQSSLLPLQHVKTLLQLDALVK 6937
QY 277 ---K--G---N---F--- 280
DB 6938 IVVARPETHNBEISDQLPISVLSVANSEPSQSFYKAGLGSVENHALNNSGGIALVFAHD 6997
QY 281 ---I---LS---AG-285
DB 6998 IREGTSMESLYGELNTRANQANLYISQGAQRDELIQVMEKSVLSYLSILAAKAGC 7057
QY 286 ---VD--- 287
DB 7058 GYLPLVPETPAARIRQILAEADVKECLTDSSMAPVIADVSRCHIMNVDTTDCSAQSGTGP 7117
QY 288 ---K---TT---I--- 291
DB 7118 QLDFKPTDIAYAVFTSGTTGPKGVLTQENILSNLEVLISKIYPVPEGSRLQACNOAFD 7177
QY 292 ---I--W---D-AH--- 296
DB 7178 VSVFIEFTWYTMCLCSASKDVMFRDFEKAINELEITHLSLTPTVAALTDPAHIPRVKF 7237
QY 297 --T-GEA---K--Q--- 302
DB 7238 LVTAGEAVTHVHWAGWAGKLYQGYCPSETTNICTVNSAVESDHVINNIGPAPENTSAPV 7297
QY 303 --Q--F--P---F---H--L---SA---P 310
DB 7298 LTQGGDDFQLVPLGGLGELCFGGQVFRGYQNMPELTESKIIINHPLYGRIYRSGDLGRLLP 7357
QY 311 ---AL---D---V---D---W--- 316
DB 7358 DGTILIQGRDORKIRGORIELGISGCLLOFPSPVQNCALIEVITADKERLMAFIPSG 7417
QY 317 ---Q---S--N--- 319
DB 7418 YSKDSYSILOPKNLEIISIAHLADNLPAVMYDVALVPVSAIPQTSQKIDKERLAS 7477
QY 320 ---N---T--FA--- 323
DB 7478 DGSALTVEDNLNAYRGADDDTSLSATEQQLASALADTLQMSQTSIGRSTSPFALGLDS 7537
QY 324 --- 323
DB 7538 VSAIRLATNLRLKEYGIDVSOILKRTIARLAPLLGGESSKOTNEPVTADCEAAVGSYL 7597
QY 324 ---S--C---S--T---DM--C-330
DB 7598 HDESUVSQLHEHGTQVSOVLPCTPLQEAMLSARDTSGSSAYRNKTLFSLHGSVDKLCACW 7657
QY 331 -I---H---V---C--K---L---G--- 337

Db 7658 EVMLQRHDLIRTIPLSTEDSRFPFVQAVLQWTLPMQECDDIPDQLSTLLDSAKAGDSI 7717
QY 338 -----H-----T-----LTK----- 440
Db 7718 VDSHPWKIQVRSSESTVYLLDMHHALYDANAMSNLLYEYQKYDQSLASAPVSKPFL 7777
QY 344 -----F-Q-----G-----H----- 348
Db 7778 NFMISTVEADALFRDLREFVPPKRTDVKSGFTITGRNLNYSKMWETFLSKHSTT 7837
QY 349 -----T-----N-----E-----V----- 352
Db 7838 MSLSTQAMMKTLAASOSYSDVCCGVNVSRSVPDGIESLVPACFTIPVRVDSLKHRS 7897
QY 353 N-----A-----I-----K-----W----- 357
Db 7898 NLGLVKALQRWNIDSLPYQLTPLRRIOQAQGTNGKRLFDSLVLQQDTTDLDSAIRLEG 7957
QY 358 -----D-----PT-----GNL-----L-----ASC-----SD 369
Db 7958 ESGWDPFCIVELAPTNESYTLSHFNRSYLDDEVSNLHQACLSAFASCIRYPSSDVS 8017
QY 370 -----DM-----TLK-----I-----WS-MK-Q----- 380
Db 8018 FIDFDADLVAGLVKPDTKHMQPVEAAKTNRSKSGSGDESMSPLEIQIRAAYSAYSSAP 8077
QY 381 -----D-----N-----C----- 383
Db 8078 EDRIERTTIYKLGLOISIAIOLANRLKDGLLVQASDVMSPCSELASAVOSRQTPV 8137
QY 384 -----V-----H----- 385
Db 8138 LDERGDFEGDKHYRGAALQSHRIATEKVASVRCPTPLQSGMLSEYTHSDGHQYFNHTF 8197
QY 386 -----D-----LQ-----OH----- 390
Db 8198 YAIEADISSKLOSAWSKVLEOHELLRTGFGVTDDEHPFVMLTYTFDFVIDLEIQASSR 8257
QY 391 -----N-----K-----E-----I----- 394
Db 8258 EGSVYSEOKASESVKNLHLPWWSLLGVSGRCLOQPSAHHAIFDAESLRLMTDLQ 8317
QY 395 -----Y-----TI-----K-----WS-----P----- 401
Db 8318 SALSNGVVPRLTIDGALHILNSQADVESQRTFWSQKLSGAPVTFPPNWTVPVRLSDTE 8377
QY 402 -----TG----- 403
Db 8378 AANVELVLYKRSKLEARCOELGVSMQSVQGAARILLSAITGESQVTFGVLSGRTPA 8437
QY 404 -----P-GTN-----N-----P-NA-N----- 412
Db 8438 TADAAFPCTITLPSVNTAVDDSQFLKDLMSYNATIQKHQFTPLTIRNYAESTSEALFD 8497
QY 413 -----M-----LA-S-----A-S----- 419
Db 8498 SLFVYQRPMDVVDSSWKIIREKASVELAVSVEMEALSDEGLGLRLTIDPAQVPHEQKG 8557
QY 420 -----P-D-----S-----T-----V-----R----- 425
Db 8558 IMLQMEVMTIAGLKFEDAINTSVMSIIPPKPIIATDFPKYLHEMTEASVKSYSRIAME 8617
QY 426 -----L-----W-----D-----VDRG----- 432
Db 8618 FVDALEDGQISSRHWTYRQDLDEANKIAHLIDRGVKPGDIATISFDKCPASAFYGIL 8677
QY 433 ----- 433
Db 8678 KAGCAFCADPTAPAAKAFILEDNARVLLTSDSIRSELRELTOCDIIDLINFNKNEL 8737
QY 434 -----C-I----- 435
Db 8738 STSSVPSGLAPSSSVYLYTSQTTGTPKGCETHDNVQLVMSFKRLFKGRWTDSSRWL 8797

QY 436 -----H-----T-----LTK----- 440
Db 8798 QFASVHFDVSVLEQFWTWIVGMRLVCAPRDILIEDIAGFLDTMQITHLDTLTPSLGRLLDP 8857
QY 441 -----H-----Q-----E-P----- 444
Db 8858 ALVPSLHKGVFTTGGBSLQDQINTWGDVGCFLPNFYGPTECTIGVTVPFCVPKEGKPSNI 8917
QY 445 -----V-Y-----S-V-----AF----- 450
Db 8918 GWQFDMVGCYVLAPGTQTPVLGAI GELCISGKLVGKYLNRPELTADCFPYLDAFCERV 8977
QY 451 -----S-----P----- 452
Db 8978 YRTGDLVRLFHGDSIDFLGRKONQVKLRQORLEIDIEIAVIRKQCIDIQDTVCIVAKHPKQ 9037
QY 453 -----D-----G----- 454
Db 9038 DKQDLIAFIGINESRKQKPELCPAESTRHLIQTABAACEERLPGYVWPTHFLPIQRIPL 9097
QY 455 -----R-----Y-----LA----- 458
Db 9098 SVNNKVEEKLRLQYADLPTTVIQTATQADSQSLSDGEQKVAQALAEKLLKIDNDLTP 9157
QY 459 -----S-G-----S-P-----D 463
Db 9158 SSNIFSLGSSISAIQFSKKLKASGFTTVQVATVTKNPTISRLTKALATSTGRSGGEIAD 9217
QY 464 -----K-----C----- 465
Db 9218 AKQVISACRQRHMGTVTRVLRCKADDIEAIACTPLQOQIIISRLASESSLYFNSFKFNA 9277
QY 466 -----W-NT-----V-H-----I- 468
Db 9278 QGVDLQKLEGAFNQALERTQILRTFETDDGVQAVRKTGLPMTWLEVYDLASVDGVF 9337
QY 469 -----W-NT----- 471
Db 9338 AKRKQKRSYNTSHLTVPFEIVIVRSGETFVSDVLDLHLYDGNSPDILMNNVSKLYNSQ 9397
QY 472 -----Q-----T----- 473
Db 9398 EADFGKFPVDCLAFGLPLRVQAKQFWLDHLPDKVSASMPPLIDNPASHDVLCASLDIL 9457
QY 474 -----G-ALV-----HSYR-GT-----G-----G-----I 486
Db 9458 NOADELRSLGVTVQALVQATWATVLRKH-YQGAIGTVVSGRSIDPDFGVENVIGPLFNTI 9516
QY 487 -----F-E-----V-C-----W----- 491
Db 9517 PFYLRCEPGDTWQTLVQRCHEFTTALPYQHTPLRDLIVKWCNKHGRPLFDALFVYQGITL 9576
QY 492 -----N-----AAG-----D----- 496
Db 9577 DNSDTNHSILKPLEDDSPADYPLSPFEAEAGGNLKIISVAAKASICNETKARELIDEPH 9636
QY 497 -----K-----VGNAS-----A-SD-----GS----- 506
Db 9637 QAFLAMNKSPENEVGASIGHTFPERLARTBDGTYKREIQARDTSDFHWSSEASVIRSEIATL 9696
QY 507 -----V-----CV-LD-----L-R-K 514
Db 9697 AGVQERADIDERTSIFEVGLDSVDVAKLSSELK 9728

RESULT 6

S02392

alpha-2-macroglobulin receptor precursor - human

N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C:Species: Homo sapiens (man)

C:Date: 14-Aug-1998 #sequence revision 14-Aug-1998 #text_change 09-Jul-2004

C:Accession: S02392; S30027; I37998; A39210; S12538

R;Herz, J.; Hamann, U.; Rogne, S.; Myklebost, O.; Gausepohl, H.; Stanley, K.K.
EMBO J. 7, 4119-4127, 1988
A;Title: Surface location and high affinity for calcium of a 500-kd liver membrane protein
A;Reference number: S02392; MUID:89210795; PMID:3266596
A;Accession: S02392
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-454 <HER>
A;Cross-references: UNIPROT:Q07954; EMBL:X13916; NID:G34338; PIDN:CAA32112.1; PID:G34339
R;Kristensen, T.
submitted to the EMBL Data Library, October 1990
A;Reference number: S30027
A;Accession: S30027
A;Molecule type: mRNA
A;Residues: 3275-3864 <KRI>
A;Cross-references: EMBL:X55077
R;Herz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.
EMBO J. 9, 1769-1776, 1990
A;Title: Proteolytic processing of the 600 kd low density lipoprotein receptor-related protein
A;Reference number: S12538; MUID:90269210; PMID:2112085
A;Contents: annotation; site of proteolytic cleavage
R;Kutt, H.; Herz, J.; Stanley, K.K.
Biochim. Biophys. Acta 1009, 229-236, 1989
A;Title: Structure of the low-density lipoprotein receptor-related protein (LRP) promotes
A;Reference number: I37998; MUID:90089395; PMID:2597675
A;Accession: I37998
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-11 <RES>
A;Cross-references: EMBL:X15424; NID:G34408; PIDN:CAA33464.1; PID:G34409
R;Strickland, D.K.; Ashcom, J.D.; Williams, S.; Burgess, W.H.; Migliorini, M.; Argaves,
J. Biol. Chem. 265, 17401-17404, 1990
A;Title: Sequence identity between the alpha2-macroglobulin receptor and low density lip
A;Reference number: A39210; MUID:91009181; PMID:1698775
A;Accession: A39210
A;Status: preliminary
A;Molecule type: protein
A;Residues: 150-166;234-238,'X',240-245,'X',247-252,'G',686-695;902-916;1096-1109;'S',17
C;Genetics:
A;Gene: GDB:LRPI; APR: LRP; LRP: A2MR
A;Cross-references: GDB:119694; OMIM:107770
A;Map position: 12q13.1-12q13.3
C;Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated protein (see PIR:A39875).
C;Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
C;Keywords: beta-hydroxyaspartic acid; beta-hydroxyaspartic acid; calcium binding; glycoprotein
F;1-13/Domain: signal sequence #status predicted <SIG>
F;20-3943/Product: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>
F;72-64/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;72-108/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;115-148/Domain: LDL receptor ligand-binding repeat homology <EG1>
F;154-188/Domain: EGF homology <EG2>
F;198-239/Domain: LDL receptor YWTD-containing repeat homology <YW01>
F;240-281/Domain: LDL receptor YWTD-containing repeat homology <YW02>
F;292-334/Domain: LDL receptor YWTD-containing repeat homology <YW03>
F;335-378/Domain: LDL receptor YWTD-containing repeat homology <YW04>
F;379-420/Domain: LDL receptor YWTD-containing repeat homology <YW05>
F;421-468/Domain: LDL receptor YWTD-containing repeat homology <YW06>
F;478-519/Domain: EGF homology <EG3>
F;571-613/Domain: LDL receptor YWTD-containing repeat homology <YW07>
F;614-659/Domain: LDL receptor YWTD-containing repeat homology <YW08>
F;660-710/Domain: LDL receptor YWTD-containing repeat homology <YW09>
F;711-752/Domain: LDL receptor YWTD-containing repeat homology <YW10>
F;753-799/Domain: LDL receptor YWTD-containing repeat homology <YW11>
F;807-842/Domain: EGF homology <EG4>
F;854-890/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;895-931/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;936-971/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F;976-1011/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F;1015-1051/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F;1062-1097/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F;1104-1140/Domain: LDL receptor ligand-binding repeat homology <LDL9>
F;1145-1182/Domain: LDL receptor ligand-binding repeat homology <LDLA>

F;1185-1221/Domain: EGF homology <EG5>
F;1227-1261/Domain: EGF homology <EG6>
F;1269-1308/Domain: LDL receptor YWTD-containing repeat homology <YW12>
F;1309-1355/Domain: LDL receptor YWTD-containing repeat homology <YW13>
F;1356-1398/Domain: LDL receptor YWTD-containing repeat homology <YW14>
F;1399-1445/Domain: LDL receptor YWTD-containing repeat homology <YW15>
F;1446-1488/Domain: LDL receptor YWTD-containing repeat homology <YW16>
F;1489-1531/Domain: LDL receptor YWTD-containing repeat homology <YW17>
F;1540-1578/Domain: EGF homology <EG7>
F;1583-1626/Domain: LDL receptor YWTD-containing repeat homology <YW18>
F;1627-1669/Domain: LDL receptor YWTD-containing repeat homology <YW19>
F;1670-1713/Domain: LDL receptor YWTD-containing repeat homology <YW20>
F;1714-1753/Domain: LDL receptor YWTD-containing repeat homology <YW21>
F;1754-1796/Domain: LDL receptor YWTD-containing repeat homology <YW22>
F;1797-1846/Domain: LDL receptor YWTD-containing repeat homology <YW23>
F;1850-1886/Domain: EGF homology <EG8>
F;1934-1976/Domain: LDL receptor YWTD-containing repeat homology <YW24>
F;1977-2019/Domain: LDL receptor YWTD-containing repeat homology <YW25>
F;2020-2063/Domain: LDL receptor YWTD-containing repeat homology <YW26>
F;2064-2105/Domain: LDL receptor YWTD-containing repeat homology <YW27>
F;2106-2151/Domain: LDL receptor YWTD-containing repeat homology <YW28>
F;2159-2194/Domain: EGF homology <EG9>
F;2193-2241/Domain: LDL receptor YWTD-containing repeat homology <YW29>
F;2253-2294/Domain: LDL receptor YWTD-containing repeat homology <YW30>
F;2344-2388/Domain: LDL receptor YWTD-containing repeat homology <YW31>
F;2389-2429/Domain: LDL receptor YWTD-containing repeat homology <YW32>
F;2430-2473/Domain: LDL receptor YWTD-containing repeat homology <YW33>
F;2482-2517/Domain: EGF homology <EG10>
F;2524-2561/Domain: LDL receptor ligand-binding repeat homology <LDLB>
F;2566-2600/Domain: LDL receptor ligand-binding repeat homology <LDLC>
F;2605-2639/Domain: LDL receptor ligand-binding repeat homology <LDLD>
F;2652-2688/Domain: LDL receptor ligand-binding repeat homology <LDLE>
F;2695-2730/Domain: LDL receptor ligand-binding repeat homology <LDLF>
F;2734-2769/Domain: LDL receptor ligand-binding repeat homology <LDLG>
F;2774-2812/Domain: LDL receptor ligand-binding repeat homology <LDLH>
F;2818-2853/Domain: LDL receptor ligand-binding repeat homology <LDLI>
F;2858-2897/Domain: LDL receptor ligand-binding repeat homology <LDLJ>
F;2904-2939/Domain: LDL receptor ligand-binding repeat homology <LDLK>
F;2944-2980/Domain: EGF homology <EG11>
F;2986-3021/Domain: EGF homology <EG12>
F;3029-3068/Domain: LDL receptor YWTD-containing repeat homology <YW34>
F;3069-3113/Domain: LDL receptor YWTD-containing repeat homology <YW35>
F;3114-3156/Domain: LDL receptor YWTD-containing repeat homology <YW36>
F;3157-3200/Domain: LDL receptor YWTD-containing repeat homology <YW37>
F;3201-3241/Domain: LDL receptor YWTD-containing repeat homology <YW38>
F;3242-3284/Domain: LDL receptor YWTD-containing repeat homology <YW39>
F;3294-3330/Domain: EGF homology <EG13>
F;3334-3369/Domain: LDL receptor ligand-binding repeat homology <LDLL>
F;3374-3408/Domain: LDL receptor ligand-binding repeat homology <LDLM>
F;3413-3448/Domain: LDL receptor ligand-binding repeat homology <LDLN>
F;3453-3489/Domain: LDL receptor ligand-binding repeat homology <LDLO>
F;3494-3531/Domain: LDL receptor ligand-binding repeat homology <LDLP>
F;3536-3570/Domain: LDL receptor ligand-binding repeat homology <LDLQ>
F;3575-3609/Domain: LDL receptor ligand-binding repeat homology <LDLR>
F;3613-3647/Domain: LDL receptor ligand-binding repeat homology <LDLS>
F;3654-3690/Domain: LDL receptor ligand-binding repeat homology <LDLT>
F;3695-3731/Domain: LDL receptor ligand-binding repeat homology <LDLU>
F;3741-3776/Domain: LDL receptor ligand-binding repeat homology <LDLV>
F;3785-3822/Domain: EGF homology <EG14>
F;3828-3860/Domain: EGF homology <EG15>
F;3868-3911/Domain: LDL receptor YWTD-containing repeat homology <YW40>
F;3913-3959/Domain: LDL receptor YWTD-containing repeat homology <YW41>
F;3944-4000/Domain: alpha-2-macroglobulin receptor 85K chain #status predicted <EX>
F;3944-4000/Domain: 85K chain extracellular #status predicted <EX>
F;3970-4012/Domain: LDL receptor YWTD-containing repeat homology <YW42>
F;4013-4056/Domain: LDL receptor YWTD-containing repeat homology <YW43>
F;4057-4099/Domain: LDL receptor YWTD-containing repeat homology <YW44>
F;4100-4142/Domain: LDL receptor YWTD-containing repeat homology <YW45>
F;4151-4182/Domain: EGF homology <EG16>
F;4200-4231/Domain: EGF homology <EG17>
F;4236-4267/Domain: EGF homology <EG18>
F;4272-4303/Domain: EGF homology <EG19>
F;4308-4339/Domain: EGF homology <EG20>

<85K>

F;4344-4374/Domain: EGF homology <EG21>	987	GRGININWRCDNDNDGDSDEAGCSHSCSTQFKCNSGRICPEHWTCDGNDGDSDE	1046
F;4377-4408/Domain: EGF homology <EG22>			
F;4421-4444/Domain: transmembrane #status predicted <TMM>	142	--AN--N-----HTD-----M-----M-E-----V-D- 152	
F;4445-4544/Domain: intracellular #status predicted <INT>			
F;166,2998/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted	1047	THANCTNOATRPDGGCHTDFQCRLDGLCIPLRWRCDGTDGDCMDSDEKSCGVTWVCDP 1106	
F;2958/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted			
F;4075,4125,4278/Binding site: carbohydrate (Asn) (covalent) #status predicted	153	--G--D-----V-----E-I--PP-----NKAV-----V- 164	
Query Match			
Best Local Similarity 70.5%; Score 2580.5; DB 1; Length 4544;			
Matches 434; Conservative 61; Mismatches 12; Indels 3805; Gaps 385;			
1 MS-----I-----SSDE-----V-----N-F-L----- 11			
88 MSRLNGVQDCMGDSDEGPHCRLEQNCRLGQCQHCVCPTLDGFTCYCNSSFOLQDGKT 147			
12 -----VY-----R-----Y-LQ-----E----- 18			
148 CKDFDECSVYTCSQLCTNTDGSFICGVBEGLLQPDNRSCAKNEPVRPPVLIANSQ 207			
19 -----SG-----FS-----H--SA-----F 26			
208 NILATYLSGAQVSTIPTSTROTMTAMDFSYANETVCWVHVGDSSAAQTOLKCARMPGLKGF 267			
27 -----TFGIK-S-HISO-----S-N-----I--NG----- 41			
268 VDEHTINISLHHVEQMAIDMTGTFYFVDDIDDRIFVCNRNGDTCVTLLDLELYNPKG 327			
42 -AL-----VP-----P-----A-A- 48			
328 IALDPAMGVFFTDYQIPKVERCDMDGQNRKTLVDSKIVFPHGITLIDLVRSLVYWADAY 387			
49 --LI-----SIIQK-L-Q-Y--V-E-----A-EVS-I-----N 67			
388 LDVIEWVDEGKGRQTIIO-GILIEHLYGLTVFENLYATNSDANAQOKTSVIRVNFN 446			
68 --E-----D-GTL-----P-----A-A- 72			
447 STEYQVTVTRVDKGGALHIYHQRPVRSHACENDQYKPGCGSDICLLANSHKARTCR 506			
73 --F-----DG-----R-----PIE-----SL-- 81			
507 RSGFSLGSDGSKCKPEHELFLVYKGRPGIIRGMDMGAKVDPDEHMIPIENLMNPRALDF 566			
82 -----S-LI-----D-----D-----AV--M-----PD-- 90			
567 HAETGFIYFADTTSYLIGROKIDGTERETILKDGIIHNVGVAVDWMDNLYWTDDGPKXT 626			
91 V-V-----QTR-----QO----- 97			
627 ISVARLEKAAQTRKTLLEGKQTHPRALVVDPLNGMWYWTWDEEDPKDSRRGRLEAAMDG 686			
98 -----ISVARLEKAAQTRKTLLEGKQTHPRALVVDPLNGMWYWTWDEEDPKDSRRGRLEAAMDG 102			
687 SHRDI FVTSKTLWPNGLSLDIPAGRLYWDADFYDRIETILLNGTDKIVYEGPELNHAF 746			
103 -----L-----L-----AQQAA----- 109			
747 GLCHGNLYFWTEYRSGSVYRLRGVGGAPPTVTLRSEPPIFEIRMYDAQOQQVGTNK 806			
110 -----AA-----A-A-----A-----A-AAQ- 120			
807 CRVNGGCSLCLATPGSRQCAEAOVDLDADGVTCLANPSYVPPQCPQGEFACANSRC 866			
121 -Q-----G-----S-----A-----K-----N-G-----E 128			
867 IQERWKCDGNDCLDNDSEAPALCHOHTCPSDRFKCENNRNRCIPNRWLCDGNDGNSDE 926			
129 -N-T--A-----N-----G-----E-----E--N 136			
927 SNATCSARTCPPNQFSCASGRCIPISWTCDDDDGDRDESASCAVPTCFPLTQTCCNN 986			
137 G-----A-----H-----T-----I----- 141			
987	Db	GRGININWRCDNDNDGDSDEAGCSHSCSTQFKCNSGRICPEHWTCDGNDGDSDE	1046
142	QY	--AN--N-----HTD-----M-----M-E-----V-D- 152	
1047	Db	THANCTNOATRPDGGCHTDFQCRLDGLCIPLRWRCDGTDGDCMDSDEKSCGVTWVCDP 1106	
153	QY	--G--D-----V-----E-I--PP-----NKAV-----V- 164	
1107	Db	SVKFGCKDSARCISKAWVCDDGNDGNDCEENCSLACRPPSHPCANNSTSVCLPDPDKJC 1166	
165	QY	-----L--R-----G--H-----E--S-----E-----V-- 172	
1167	Db	DGNDGCGSDGEGELCQCQLNNGGCSHNCVAPGSGIVCSCPLGMLGPDNHTCOIQSY 1226	
173	QY	-----F-I-CA--W-----N-----N----- 178	
1227	Db	CAHKLKCSQCKDQNFVKSCVEGWVLEPFDGSCRSRDPFKPFIIFSRRHEIRRIDLHK 1286	
179	QY	-----P--V-----S-----D-----LL--A--S----- 186	
1287	Db	GDYSLVPLGRNTIALDFHLSQALYWTVDVEDKIYRGKLLDNGALTSFEVVIQYGLATP 1346	
187	QY	-----G-----S-----GD-----S-----T-- 192	
1347	Db	EGLAVDWIAGNIYVWESNLDQIEVAKLDGTLRTLLAGDIEHPRAIALDPRDGLFWTDW 1406	
193	QY	-A--RI-----W-N----- 197	
1407	Db	DASLPRIEASMSGAGRTVHRETSGGWPNGITVDYLEKRLIMWDARSDAIYARYDGS 1466	
198	QY	-----LS-----E-----NST-----SG-----S-TO--L 209	
1467	Db	GMEVLRGHEFLSHPPAVTLYGGEVYTWDTWTH-TLAKANKWTGHNVTVQRTNTPFDL 1525	
210	QY	-V-----L--R-----H-----C----- 214	
1526	Db	QVYHPSRQPMAPNCPCEANGGQPCSHLCILINYNRTVSCAPHLMLKHKDNTCYEFKKFL 1585	
215	QY	-----I-----RE----- 217	
1586	Db	LYARQMEIRGVLDAPYNYIISFTVPDIDNVTLVDYDAREQVYWSVDVTRTOAIKRAFIN 1645	
218	QY	G-G-----QDVPSNK--D-V-----TS-----LD----- 232	
1646	Db	GTGVEITVSADLP-NAHGLAVDWVSRNLFWTSDYTNKKQINVARLDGSPKNAVQGLEQP 1704	
233	QY	-----W--N--S-----EG--TLL--A--TG-----S--Y-- 245	
1705	Db	HGLVHPLRGKLYWTGDNISMANMDGSRNLTLLFSGQKGPVGLAIDFPESKLYWISSGNH 1764	
246	QY	-----D-GF--A-R-----I-----W-----T--K-DG----- 256	
1765	Db	TINRCLDGSGLVIDAMRSQKATALAIMGKLWADQVSEKMGTCADGSGSVLR 1824	
257	QY	N--LA-----S-----T-----L--GQ- 264	
1825	Db	NSTTLVHMVYDESIOLDHKGTPNCSVANGDCSQLCLPTSETTRSCMCTAGTSLRSQQ 1884	
265	QY	-----HKG-----P-----I-F-A-----L- 272	
1885	Db	ACEGVGSFLLYSVHEGIRGLPDPNDKDALVPVSGTSLAVGIDFHAENDTIYVWDMGLS 1944	
273	QY	-----K--W--N--K-----KGN-----F-I--L-----S 283	
1945	Db	TISRAKQDTWREDVVVNTNGIRVEGIADWIAWNIYTDGDFVIEVARLNGSFRYVVIS 2004	
284	QY	AGYDK-----T-----T-----I-----I-- 292	
2005	Db	QSLDKPRATVHPKGYLFTWENGQYPIERSRLDGTERRVVLVNVVSIWPNGLISVDYQDG 2064	
293	QY	--W-DA-----H----- 296	

Db 2065 KLYWCARDTKIERIBLETGENREVLSSNNMDMFSVSFEDFIYWSDRTHANGSIKRG 2124
QY 297 ---TG--- 298
Db 2125 KONATSVPLRTIGVQLDKIVFNDRDRQKGTWCAVANGCCOQLCLYRGRGORACAH 2184
QY 299 ---E-A---K---Q---Q---F--- 304
Db 2185 GMLAEDGASCREVAGYLLYSERTILKSIHLSDBRNLANAPVQPPEDPEHMKVIALAFDYR 2244
QY 305 ---P---F---H---S---A--- 309
Db 2245 AGTSPGTNRIFFSDIHFGNIQINDGGRRIITIVENVGSVEGLAYHRGMDTLTYSTT 2304
QY 310 ---P-AL---D---V-D---W-Q---S 318
Db 2305 SITITHTVDTQTRCAFERETIVTMSGDDHPRAVFLDECQNLMPWTWNEQHSIMRAALS 2364
QY 319 ---N---N---T--- 321
Db 2365 GANVLTLIEKDITPTNGLAIDHRAEKLYPSDATLTKIERCEYDGSYHYILKSEPVHPFG 2424
QY 322 ---F---F---A---SC-S--- 326
Db 2425 LAVYGEHIFWTDWVRRAVORANKHVGSNMKLLRVDIPOQPMGIIVANDTNSCELSPCRI 2484
QY 327 ---T---DM---CI--- 331
Db 2485 NNGGQDCLLTHQGHVNSCRGRILQDLDLTCAVNSSCRAQDEPECANGECINFSLTC 2544
QY 332 ---H---V---C-K--- 335
Db 2545 DGVPHCKDKSDEKPSYCNRRCKTFRQCSNGRCVSNMLWCNADCGDGSDEIPCNKTA 2604
QY 336 ---IG---Q--- 338
Db 2605 CGVGEPRCRDGTICIGNSSRCNQFVDCEDASDEMNCSATDCSSYFRGLGVLFQPCERTS 2664
QY 339 ---D---RP---I---K--- 343
Db 2665 LCVPASWCDGANDCGDYSDERDCPVGRPCPLNYFACPSGRCI PMSWTCKDEDDCEHG 2724
QY 344 ---T---F---Q---G---H---T 349
Db 2725 EDETHCNKFCSEAFQCNHRCISKQWLCDDGSDGDEAAHCRGKTCGPSSFCPT 2784
QY 350 ---N---EV---N-A-I-K--- 356
Db 2785 HVCVPERWLCDDKOCADGADESIAAGCLYNSTCDDREFMCONRQCIPKHFVCDHRRDCA 2844
QY 357 ---W---D---P-T--- 360
Db 2845 DGSDESPECEYPTCGSPBPCANGRCLSRQWECGENDCHDQSDAPKNPHCTSPCHKC 2904
QY 361 ---G---N---L--- 363
Db 2905 NASSQFLCSSGRCAEALLCNGQDDCGSDSDRGCHINECLSRKLSGCSDCEDLKIGFK 2964
QY 364 ---L---A---S---CS---D--- 369
Db 2965 CRCPGFLKDDGRTCADVDECSSTPPCSQRCINTHGSYKCLCEVYAPRGDPHSCAV 3024
QY 370 -D---M---T-I-K---I-W---SM--- 378
Db 3025 TDEEPLIFANRYLRKLNLDGNSYTLTKQGLNNAVALDFDYREQMIYTDVTTOGSMIR 3084
QY 379 -K---Q---D---N---C--- 383
Db 3085 RMHLNGSNVQVHLRHTGLSPDGLAVDWVGNLYWCKDGRDTIEVSKLNGAYRTVLVSSGL 3144
QY 384 ---V---H--- 385
Db 3145 REPRALVVDQNGYLYWTDWGDHSLIGRIGMDSSRSVIVDTKI TWPNGLTLDYVTERIY 3204

QY 386 ---D---L---Q---Q---H---N---K--- 392
Db 3205 WADAREDIYEFASLDGSRHVVLSQDI PHIFALTLPEDVYVWTDWETKSNRAHKTGTN 3264
QY 393 E---I--- 395
Db 3265 KTLILSTLHRPMDLHVHFHALRQPDVFNHPCVKNNGGCSNLCLLSPGGGHKACPTNPLYG 3324
QY 396 ---T---K-W---S---P---TG 403
Db 3325 SDGRTCVSNCTASQFVCKNDKCI PFWMKCDTDDCGHSDDEPPDCPEFKRPGQFCSTG 3384
QY 404 ---P---G---TN---N--- 408
Db 3385 ICTNPAFICDGDNDQDNDSEANCDIHVCLPSQFKCTNTNRCIPGIFRCNGQDNCGDGED 3444
QY 409 ---PN---ANL-M--- 414
Db 3445 ERDCPEVTCAPNQFQCSITKRCIPRVWVCDRDNDCVDSDEPANCTQMTGCVDEPRCKDS 3504
QY 415 ---L-A--- 416
Db 3505 GRICIPARWKDGEDDCGSDGDEPKCECDERTCEPYQFRCKNNRCVPRWQCYDNDCGDN 3564
QY 417 ---S---A-S---F--- 420
Db 3565 SDEBSCTPRPCESEBFCANGRCIAGRWCDDGHDCA DGSDEKDCPTPCMDQFQCKSGH 3624
QY 421 ---D-S---T-VR---L-W---D 428
Db 3625 CIPLRWCDADADCMGDSDEBACGTGVRTCLPDEFCQNTLCKPLAWKCDGEDDCGDSND 3684
QY 429 ---V---DRGIC-I---HTLTKH- 441
Db 3685 ENPECAFCVCPNRPFRCKNDR-VCLMTGROCDGTNCGDGTDEBDCBPPTAHT-T-HC 3741
QY 442 ---Q---E---P---V--- 445
Db 3742 KDKKFLCRNQRLSSLRNMFDDCGDGSDEDCSIDPKLTSCATNASICGDEARCVRT 3801
QY 446 ---Y---S---V---A-F--- 450
Db 3802 EKAAVCACRSQFHTVPGQPCQDINECLRFGTCSQLCNNTKGHLCS CARNFMKHTNCK 3861
QY 451 ---S---P---D---GR--- 455
Db 3862 ABGSEYQVLYIADDDNEIRSLFPFPHSAYEQAFQGDSEVRIDAMDVHVKAGRVYTNWHT 3921
QY 456 ---Y-L-A-S-G-S-F-D- 463
Db 3922 GTISTRSLPPAAPPTTSNRRHQIDRGVTHLNLISGLKMPRGIAIDWVAGNVYVWTDSDGRV 3981
QY 464 ---K-C---V-H-I---W-N---T--- 471
Db 3982 IEVAQMKGNRKTLLISGMIDEPHAIWVDPLRGTYMSDGNHNPKIETAAMDGLRTELQV 4041
QY 472 ---Q-TGAL-V-HS-Y---R-GT---G--- 484
Db 4042 DNIQWPTG-LAVDYHNERLYWADAKLSVITGSRNLNGTDPIVAADSKRGLSHPSIDVFED 4100
QY 485 ---G---I-F---BV---C-W--- 491
Db 4101 YIYGVTYINNRFKIHKGHSPLVNL TGGLSHASDVVLVHQHKQPEVTPCORKKCEWLC 4160
QY 492 ---N---A--- 494
Db 4161 LLSPSPVCTCPNGKRLDNGTCVPVSPPTPPDAPRGTCNLQC FNGGSCFLNARQPKC 4220
QY 495 ---GDK---V-G-A--- 500
Db 4221 RQPRYTGDKCELDQCEWHCRNGGTCAASP SGMPTCRPTGTGPKCTQQVCAGYCANNS 4280

Qy 501 -----S-----ASDGS----- 506
Db 4281 TCTVQGNQPCRLPGFLGRCQYRQSGYCNFGTCQMAADGRCRCRTAYPEGSRC 4340
Qy 507 V-----CVL-----DL-----R-----K 514
Db 4341 VNKSRCLGACVVKQSGDVTCTDGRVAPSLCTCVGHGNSGSGTMSK 4392

RESULT 7
T13931
projectin - fruit fly (Drosophila melanogaster) (fragment)
C;Species: Drosophila melanogaster
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T13931
R;Daley, J.; Southgate, R.; Ayme-Southgate, A.
J. Mol. Biol. 279, 201-210, 1998
A;Title: Structure of the Drosophila projectin protein: isoforms and implication for pro
A;Reference number: 217815; MUID:98300339; PMID:9636710
A;Accession: T13931
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-6658 <DAL>
A;Cross-references: UNIPROT:O76281; EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AACT
C;Genetics:
A;Gene: projectin
A;Cross-references: FlyBase:FBgn0005666
A;Map position: 4
A;Note: intron positions not resolved (incomplete sequence)
C;Keywords: muscle

Query Match 70.5%; Score 2580.2; DB 2; Length 6658;
Best Local Similarity 8.8%; Pred. No. 3e-52; Mismatches 10; Indels 4578; Gaps 400;
Matches 446; Conservative 55;

Qy 1 M-S-----IS-----S-----D-----EVN----- 9
Db 674 LPSSEIKNIPYNTKISIIETVRKHTGIYKIIAVNEHGQDEATVEVNILAPPSKPRGLD 733
Qy 10 -----F-----L-VY-----RY----- 19
Db 734 VKDVTDSCKFKWKPKBEDDGGKPIQAYQVEKFKDQKQWVPLGRTSANDTFDVKGLQE- 792
Qy 20 G-----F-----S-----H-----S 24
Db 793 GHEVQFRVKAINEEGESDPLSDSIIAKNPYDAASKPGTGNIVDYNEHVMVKLWEAPRS 852
Qy 25 -----A-----F-----T-----F-----GI--K 31
Db 853 DGGAPISGIYIIKKDKFSPIDWIELSTNTSVPATVEGLVEGNIYQFRVAVNKAQFSDP 912
Qy 32 S-----H-----I-----S-----S 35
Db 913 SDATEPHAKRNLKPYINRDKMKPIKVRAGQPVKFDVDVKGEPAPSLTWFLKETELTST 972
Qy 36 -----OS-----NING-----A-----L-V- 44
Db 973 GQVRLNIDYNTKLTLLDTRKQSGQYKLAENINGVDEAVVEVILDKPKSPEGPLEVS 1032
Qy 45 -----P-----P-AAL-I-----SI-----IQ-KGL-----Q 58
Db 1033 DIHKEGCKLWRKPKDGGIPITGVYIEKMDTATGKWPAGSVDPKDYIEIKGLDPNHR 1092
Qy 59 Y-----V-----E-----A-----E-V----- 64
Db 1093 YQFRVAVNEEGESEPLETESATAKNPFDVSPAPGLPELEDWDEHHVKLWEPPIRDGG 1152
Qy 65 S-I-N-----E-----D-G-----T-----L----- 72
Db 1153 SPITNVIIEVMDKDPGEFVKAVETDSPVKGVVKKLEEGQYKFRVAVNKAQSPDSEQ 1212
Qy 73 -----F-----D-----G-----R-----P-IE-----S 80

Db 1213 TNHHVAKPRFLKPHIDRVNLKPVIVKTLGLSILDIRGEPAPKVEFFNNSSVTSDEHS 1272
Qy 81 -----L-----S-LI-----D-A-----V----- 87
Db 1273 VKIDNDVYNTKFFVMRAORSQSGYIIKATNEVGEDEALEVTVLGKPGKPKGLQVNDI 1332
Qy 88 -----MPD-----V-----V-----Q----- 93
Db 1333 TKHSCKLKWEKPDGDDGSPIDYIEIEKLDPHTCQWLPCGKSTPEAKVIGLHKGKAYKPR 1392
Qy 94 -----T-----R-----Q 96
Db 1393 VRAVKEGESEDLTEKPIIAKNPYDEPDRPGKPEPTNWDKDFDLAWDPKNDGGAPIQ 1452
Qy 97 -----Q-----A-----Y-R-----D-K-- 102
Db 1453 KYVIQMRDKSGRAVDSATVPDGKNGTGTGVBEGHEYEFRIIVAVNKAQSPDSDVSKSV 1512
Qy 103 LA-----Q-----Q----- 106
Db 1513 IAKPRFLKPHIDRKNLQKKIMRSGQMLHIDALIKABPPAKVTWYNTKTEIKTSDHIKIE 1572
Qy 107 -----QA-----A-----A-----A----- 111
Db 1573 EDYKTTFIMPVKRADRGYIIVTAKNDSGSDTVEVEVLCKPSKPKGLAVSNVTAETL 1632
Qy 112 -----A-----A-----A----- 113
Db 1633 HLKWEKPEDDGGDPIBOYLVERMDTETGRWVPLTTKTTEADVTGLTEGKEYLFRVKA 1692
Qy 114 -----A-----AA-----A-----A----- 119
Db 1693 SEGESEPLVTIDPTKAKNPFDAADTPGKQIIVDWSNGHCDLKWRAPEDDGGASITGYVE 1752
Qy 120 -----Q-----Q-----Q-----G-S-----AK-- 125
Db 1753 RKDPNTGKQKALETSTPDKARVNDLIAGNKYQFRIMAVNKAQSKPSPSQMTAKDR 1812
Qy 126 -----N-----GE-----N----- 129
Db 1813 FAPPKIDRNIKIDITSKAGQIRFDIKVSGEPPATKVLHKNKARLENDNSNYNIDMESVR 1872
Qy 130 -----T-A-N-----G--E-----E----- 135
Db 1873 TKLTVPISKRFHSGKYTLKAENESGRDEASFEVIVLDKPGPEGLRVTDVHKEGCKLKW 1932
Qy 136 N-----G-AH-TI-----A--NN-----H-----T----- 146
Db 1933 NAPLDDGGLPIDHYIIIEKMDVESGRWLPSPGRFKESPAELNNLEPSHEYKFRVLA 1992
Qy 147 -----D--M-M-E-----VD----- 152
Db 1993 SBPLTGEQSVIAKNPFDEPGKPGTPEAVDWDKHDVLDVWRPPINDGSGSPITGYVVEKREK 2052
Qy 153 G-D--V-----E-IP-----P--N-----KAV----- 163
Db 2053 GTDKWIKGTEITPCLGEECKATVPTLNCEVFEVFRVKAINAAGPGEPSDASKPIITKPR 2112
Qy 164 -----VL-----R-----G-----H-E----- 169
Db 2113 KLAPTILDPKIRTYNFKSGSPIFLDINISGEPADVTNQNKNKSVQTTFSHIENLPY 2172
Qy 170 -S-----E-----V-F-----I-----C----- 175
Db 2173 NTKYINNPERKDTGLYKISAHNFYQQDQVEFQIINIITKPGKPGGLEVEVHKGCKL 2232
Qy 176 -----S-----D-L-----L-----AS--GS----- 180
Db 2233 WKKPKDDGGEVSVYVEKFDPTDGLVPGVGRSDGPEYNVDGLVPGHDKYKFRVKA 2292
Qy 181 -----S-----D-L-----L-----AS--GS----- 188
Db 2293 ESEPLETLGSIIAKOPFSVPTKPGVPEPTDWTANKVELAWPEPASDGGSPIOGYIVEK 2352

QY 189 -----G-----D-S--TA----- 193
Db 2353 KYSPLWEKALETNSPTTATVQGLBEGNEYQFRVVALNKGNLSEPSDKIFTAKRYIA 2412
QY 194 -----R-----I-----W-----N-----L- 198
Db 2413 PKIDRNLRLNITLSSGALKLDANITGQAPKPVWKLNYHLQSGKNVTIETPDYTKLV 2472
QY 199 -----S-E-----N-----S-----T-----S 204
Db 2473 IRPQRTDSEYLVATNTSGKDSVLNVVVITDKPSPPNGPLOISDVHKEGCHLWKRPS 2532
QY 205 -----G-----ST-Q-----L-----V-----L 211
Db 2533 HDGGTPIEYQIDKLEPETGCMIPSCRSAPQVDVTCLSPGNEYKFRVSAVNAEGESQPL 2592
QY 212 -----R-----H-----C-I----- 215
Db 2593 VGDESIVARNPFDEPGKPNLKATDMDKHVDLAWTPPVIDGGSPISCVIIEKQKYKGM 2652
QY 216 -R-----EG----- 218
Db 2653 ERALVPAOCKATIPDLVEGQYKFRVSAVNAAGTGPESDSTPPIIAKARNKPIIDRS 2712
QY 219 -----GQ-----D-V-----P-----S-----N-K----- 226
Db 2713 SLVEVRIKAGQSFTFDCKVSGEPAPQTKMLKKKEVYKONKVTNVDTYTKLVNSATR 2772
QY 227 -----D-V-----VT-----S----- 230
Db 2773 SDSGIVTVFAENANGESADVKTVIDKPAPPNGPLKVDEINSECTLHWNPPDDGGOP 2832
QY 231 -----LD-----W-----N-----SE----- 236
Db 2833 IDNVVVGKLDETTGRWMTAGETDGPVVALKVGGLTPECHKYKFRVRAKNRQGTSEPLTAA 2892
QY 237 -----G-TL-----L-----A-----TG----- 243
Db 2893 AIIAKNPFDPVTPKPTTKDFKDFVDLEWTPREADGGSPITGYVVEKDKFSPWEKC 2952
QY 244 -S-----Y----- 245
Db 2953 ABISDDITIAHVLDLIEGLKYEPRVAVNAKAGCPSPDATETHVARPKNTPPKIDRNEMS 3012
QY 246 -----D-----G-----F----- 248
Db 3013 DIKAGNVFEDVPVTGPEPLPSKDWTHEGNMIINTDRVKISNFDRTKIRILSATSDTG 3072
QY 249 -----AR-----I-----W-----T----- 253
Db 3073 VYTLTARNINGTRHNKVTILDAPSPPEPALRNGDVSKNSIVLRWRPPKDDGGSBIHY 3132
QY 254 -K-D-----G-----NL-----A-----S-T----- 261
Db 3133 VVEKMDNEAMRWVPVGDCTDTEIRADNLNENHDYSFRVAVNAKQOSQPLTTSQPIATKD 3192
QY 262 -L-----GQ-----H-----K-G-PI-----P-----A----- 271
Db 3193 PVSHDPKQGPQATDNGKHFDVLEWSTPKRDKGAPISVYIIEKRPFGQWERAUVVLGDN 3252
QY 272 -----L-----K----- 273
Db 3253 CKAHVPELTNGGEYFRVAVNRRGGSDPSDSTIICKPRFLAPFEDKSLNDIIVHAG 3312
QY 274 -----W-N-KK-----G-N-F-I-L-SA----- 284
Db 3313 NGLGWTLPTEASPRPLITWLYNCKEBSNRSRGSGLFQNELTEIYVSSLSRSABGRVTLIL 3372
QY 285 -G-----V-----D-K-T-----T-----II----- 292
Db 3373 KNEHGSFASAHATVLDPRPPKGPLDITKITRDGCHLTWNVPDDGGSPILHYIIEKMD 3432

QY 293 -----W-DA-----H-T-----GE-----AK----- 301
Db 3433 LSRSTWSDAGSMTHIVHDVTRLVHRKEYLFRVKAVNAIGESDPLEAVNTIIIAKNEFDEPD 3492
QY 302 -----Q----- 302
Db 3493 APFKLIITDWRDHDLDLOWAVPKSGGAPISEYIIQKKEKSPYTWNRVHPSNKNNTTI 3552
QY 303 -----Q-F-----P-----PH----- 307
Db 3553 PELTEQGEYFRVAVNAQAGQSEPSDMMKRPYLPKPIITPLNEVRIKGLIFHTD 3612
QY 308 ----- 307
Db 3613 IHFIGEPAPEATWTLNSNPLLSNDRSTIITSIGHSVVHTVNCQSDSGIVHLLLRSSGI 3672
QY 308 ----- 307
Db 3673 DEGSFELVLDPRGPPGPMYEEITANSVTISWKPPKONGGSEISSYVIEKRDLTHTGG 3732
QY 308 -----SA-----P-----A-----L-D-----V----- 314
Db 3733 WVPANVYSAKYNHAVVPRLLEGTMYELRVMAENLOGRSDPLTSDQPVVAKSQYTVPGAP 3792
QY 315 -----D-----W-Q-SN-----NT----- 321
Db 3793 GKPELTDSKNHITIKWKQPISSNGGSPIIIGYDIERRDVNTGRWIKINGQVPVTAEQDDR 3852
QY 322 -----F----- 322
Db 3853 VTSNHOYQYRISAVNAAGKGTSEPSAIFNARPLREKPRFYFDGLICKRIKVRAGEPVNL 3912
QY 323 -A-----SC----- 325
Db 3913 NIPISGAPTIEWKKGDLKBEGKRISEYETNSERTLFRIDDSNRDRSGKYTVTAANEFG 3972
QY 326 -----S-T-DM----- 329
Db 3973 KDTADIEVIVDKSPPEGFLSYTETAPDHSILHWYSPKDDGSDITGYIIEFTEFGVDD 4032
QY 330 -C----- 330
Db 4033 WKPVPGTCNTNFTVKNLVEGKKYVFRIRAEINYGASEALEGKPLVKSPDFPGAPSQP 4092
QY 331 -I-----H-----V-----C----- 334
Db 4093 TISAYTPNSANLEWHPPDDCGGKPIITGYIVERRERGGEWIKCNYPTPTNTSYVSNLRDG 4152
QY 335 -K-L-----G-----Q-----DR-----P----- 341
Db 4153 ARYEFRLAVNEAGPGHPSKPSDPMPTAEHQRYRPPPEPPKPDRIITNGVTLNWRPRTD 4212
QY 342 -IK-----T-----F----- 345
Db 4213 GKSRINGYYVEMRPNKNGDKWKTVDIPINSTVYVTPSLKEGEYSFRVVAENEGVRSDDS 4272
QY 346 -----Q-----G----- 347
Db 4273 KPSQPTIIEQPNKPCMLGKVRDIVCRAGDDFHSIHVPYLAFFPKPAFWYSNDNMLDDNN 4332
QY 348 -H----- 348
Db 4333 RVHKHLTDDAASVVVKNKSRDGGQYRLOFKOTSGFDATINVRVLDPRSPPTLRADVF 4392
QY 349 -----T-----N-E----- 351
Db 4393 SGDSLTYWNPNDGGSIAQNYIIEKKEARSSTWSKVSFCTVFLVRIRNLVLNKEYDF 4452
QY 352 -V-----N-----A-----I-K-----W-----D-----P----- 359
Db 4453 RVIAENKQSGDPANTSEPIARHPDIPNTPGIPIHGIDSTEDSTIATMTKPKHGGSP 4512
QY 360 TG-----NL-----L-----A-----S----- 366

Db 4513 TGYIIKELLSDDKWTKAVHALCPDLCKIENLNIENAEYERVAAVNAQSGSSDL 4572
QY 367 -C-SD-DMT-LKI-W-376
Db 4573 IFCRRPHAPKITSDLSIPDMTVIAGDEFRITVPHASPRPTASNGLEVIPGERIKF 4632
QY 377 -SM-KOD-N-CVH-D-L-Q-388
Db 4633 DSNYASYNKSARDETSYITLTNNKSGDTASC-HVTVDRLPPQGLNAYDITP 4691
QY 389 -Q-390
Db 4692 DTCTLAWKTPLDDGSPITNYVVEKLDNSGWSKISSFVRNTHVDVNGLEPHYKYNFRVR 4751
QY 391 -N-K-E-IYT-396
Db 4752 AENQYGLSDPLDIIEPMVAKHQFTVPDEPGQKVIDWDSGNVTLIWTRPLSDGSRIOGY 4811
QY 397 -IK-W-SPT-402
Db 4812 QIEYRDILNDSSWAYDYIIKTKYQLYNLNGSEYEFRIKAKNAAGLSKPSPSLRFKL 4871
QY 403 -G-PG-T-N-NP-S-410
Db 4872 KGKFTVPSPPGAPQVTRGVKNYVDLKWEKPLRDGGSRIITGVIIERRDIGGAVVWKNDYN 4931
QY 411 -ANL-M-LA-SA-S-419
Db 4932 VLDTEYTMNLIEMGDYEFVRFAVNSAGRSEPSLCTMPIKVCYGLGKPKDWITRLQDKV 4991
QY 420 -F-D-S-TVRLW-DV-D-R-431
Db 4992 APFGKDYTLQCAASGKSPAR-WLRNGKSIQMGGRMTCDKDGVRFLHSINVQTGDDG 5050
QY 432 -G-I-C-434
Db 5051 DYTCEAMNLSLGFVNTSGYLKIGSPIINRCPSELKPEGDNSKIKIFYSGDQPLTVILKKN 5110
QY 435 -I-H-T-LT-K-H-441
Db 5111 NEVISDNDTHVKVNFDDYVAIYIANIVKSGDGPQVQIEFTNESGSGATGEFYVHTGMP 5170
QY 442 -QEP-V-Y-S-V-AF-450
Db 5171 SAPTGMGISVINKNSCMLNWRPPYDGLKVSHYVTERKDVSSPHWITVSSCKOTAFN 5230
QY 451 -SP-452
Db 5231 VOGLIENQEIYFRMAVNMENGMGPPLGRLNPIRAKDPIDPPSPPAVLKSLRSBETLCNPE 5290
QY 453 -D-GR-Y-456
Db 5291 WKPESDGGAHIQGYWIDKREVSCKMAGVNATICAANQINCINLIEGROYEPRIPAQNV 5350
QY 457 -L-A-S-GSF-D-K-CV-H-I-W-469
Db 5351 AGLSTAIVRTSSQDNRSCTGGFASIDCETATHANCIONHNAQFTCTIPGVPKPTISWYKG 5410
QY 470 -N-TQ-T-GA-475
Db 5411 AREISNGAGITCTQKVTSLPKYYDVFGEADAEYVCRAVNAKAGAKSTRATLAIMTAPKLVN 5470
QY 476 -L-VH-S-Y-R-481
Db 5471 PPRFRDTAYFDKGENVVIKIPFTGLPKPRHWWKGLNIESGGHYTVVEKHAVALIIRD 5530
QY 482 G-TG-G-I-F-E-V-C-490
Db 5531 GSHLDSGPYRITAENELGSDTAIIQVQISDRPDPPRPLIESIGTESLSLSWKPAPWDGC 5590
QY 491 -W-N-A-G-D-496

Db 5591 SDITNYVVERDDPLSSWIRVGNTRFTSMAVSGLTGPKYEDFRIADNVVYGRSDASDTST 5650
QY 497 -K-V-G-A-S-AS-503
Db 5651 LIKTESVKKKPIERKWEIDANGRLRGKADGPKDYDSYVFDIYSKFPQPPVEISQOSV 5710
QY 504 -D-GS-V-C-VL-DL-RK-514
Db 5711 YDRYDILEIGTAGFVGHRCRERSTGNIFAAKFIPIVSHSVEKDLIRRE 5759
RESULT 8
T15789
hypothetical protein C41A3.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T15789
R;Bentley, D.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C41A3.
A;Reference number: Z18404
A;Accession: T15789
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7829 <BEN>
A;Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA8311
C;Genetics:
A;Gene: CESP.C41A3.1
A;Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/2;
/2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein
F;6547-6616/Domain: acyl carrier protein homology <ACP>
F;2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted
Query Match 70.5%; Score 2578.7; DB 2; Length 7829;
Best Local Similarity 8.2%; Pred. No. 4.5e-52;
Matches 456; Conservative 45; Mismatches 11; Indels 5083; Gaps 372;
QY 1 MS--ISS-----D-----E-----V-N-F-----LV---12
Db 1 MSACLVGSAKIPNGEDGHEMAENIFLNRNIALLPVKNYLSDFREKHEPVKAALVDGI 60
QY 13 -----Y-----R-----Y-15
Db 61 EYFDDQYFGTGESEATCMPDQRMQLMGVYKGLNAGITILEMASEARVAVYTAACVQDYK 120
QY 16 -L-Q-----E-----SG---20
Db 121 DLLPPDQYMATGNSASVMCGRITYFLNSRGAAGVIGTACSSSLVAFHLARQATQSGETKL 180
QY 21 -----F-SH-----SA-F-TF-----28
Db 181 ALVCGANHVGSRFSHLYNSHWSVNGRLAADRANGFVRAESFAVAVLCSKQFAEEN 240
QY 29 -----G-----I-K-S-----H-33
Db 241 LLHCECVGSFAFNSDGTSLTAPNPIISOYEQLEAKNIDKDSVQLVTCGTGTLGQ 300
QY 34 -----I-S-----QS-----N-38
Db 301 VELTAINRFSKDIRVMSPKSSMGHGEAAGLIGVLSLYSMQHGIIPNQLHLELPSDL 360
QY 39 -----IN-----G-ALV-P-P-----A-48
Db 361 GEDKSMGFVNEEMELNRVAISSYGGGTNACAIIEKPEKPSLVQKESYAESNVLFSAKS 420
QY 49 -----L-I-----S-I-----I---QK---56
Db 421 HESLKLQIEYTFQFMAQSDSAMEDIITYNTERKTYDFRAAVFGKNEEIAKRLQGDYS 480
QY 57 ---IQ-----Y-----V-----EA---E---63
Db 481 LTNLOESTFEVFGEGNEKLWLLRLMYEKNETFTHTVDVKYCKLAETCGPFEARTALFPDF 540

QY 64 -----VS-I-----N----- 67
Db 541 KLTPLTYNSRLISSMATFELLVQYNTLPNKLKRGKGLQIFCLAVAKVITESAVQLI 600
QY 68 -----E-----D-----G-----T----- 71
Db 601 KGVBAENLTDILGDIELXSSKIPIBIOHLKSTKKILPHIISGELKETAKPNLMTFIYN 660
QY 72 -----L-----FD-----GR-----P-----I 78
Db 661 GBEILEDPVRKVQKLIQCLFACGFPDPAVKFRGRIVKTPTYSPLKQFPEVQGTAMTWI 720
QY 79 -----SIS-----L-----I-----D-----AV----- 87
Db 721 VDEQTNSSLSDAEISTVTVIVKQFLDIEDDINLLETGAVDLSLTSIEMVEAFGTAVNQT 780
QY 88 MP-----D-----D-----D-----D-----Q----- 93
Db 781 MPFDLLEAYPTILNIVDFLKLTVTVPTVKATTSIHKTSSELSTSDINVIACDYQFAGVE 840
QY 94 -----T-----R-----Q-----Q-----A----- 96
Db 841 GEKELWDTLLTSLTKGIDIRKKQCEGDAGLEVGLLKQDISMFDNSFFAIAXDBAEFL 900
QY 97 -----AY-----A-----A-----A-----A----- 99
Db 901 DPQHRLLNAAYNALXSKSLTSPDADLFLAISAHSEYRALAEKHINELDERLWMTVHS 960
QY 100 -----R-----D-----KLA-----Q-----Q----- 106
Db 961 MVAGRLAVLMGIRGAMIVDTTCSSVATALEMAVKSIREGKFAIVATSQLIQSSKWLYS 1020
QY 107 -----Q-----Q-----A-----A-----A-----A----- 111
Db 1021 LKTLDDHSTNSFSDGSGFCRDGCVIILKTAEGDSAVIKISSAKSHHCGAVMTPV 1080
QY 112 -----A-----A-----A-----A-----A-----A----- 119
Db 1081 SSISQLEAEAGSFYVEGHGTATSDAGSAESMAYQKLGSELINSSVKAQFGHCEVASGLI 1140
QY 120 Q-----Q-----Q-----Q-----Q-----Q-----G----- 122
Db 1141 QLMKVSSIGKHGIIPSVHNILPSEHNRNENIRLPFAVEEKQIDRSIAIVSFGITGKTIV 1200
QY 123 -----SAK-----S-----S-----S-----S----- 125
Db 1201 VTERVSQLNVONIQCNYLLPVSATKQGLKXKACLSLIEMIDNSCESLYDISTTLOKQKT 1260
QY 126 N-----G-----E-NT--AN-----G-----E-----E----- 134
Db 1261 NFKWRTAVVGSADVVVLKQKFLTSEHNTSLTNWHISTSHSISGCGSTFFHNIPEFEDHY 1320
QY 135 -----E-----NG-----A-----H-T-----IA-N----- 143
Db 1321 SMFCHRLRPEPHSNTNSIVHLLAVVALIRVILKHLKLTNSFVAGGPNLSLIVLAAVDAAPS 1380
QY 144 ---N---HT---D---MM-----E-----V----- 151
Db 1381 HYLNDLLHAFANDVVMKRIARDVTISINVKLNLNAGEPITTARQAVEATIDQKVXV 1440
QY 152 -----D-----D-----G-D-----V-----E----- 156
Db 1441 RLPETTLILSPRAYEFASQLETIQDYKYLIGEKTQGGQVDFAGIFGTPIKLIDLPEYP 1500
QY 157 -----IP---P-N-----KA-----VVL-----R- 166
Db 1501 FNRKSWFLPIPSVPSNEKGEKPLIPKSYEFLLKSQWQHVQNVVDSKIVLPGATSIRL 1560
QY 167 -----G-----G-----G-----G-----G----- 167
Db 1561 VHQLNGKPTVELSNIDFLNKTIPSEAPSVVKIEEQDGLKLVFGETDAISFKLTQLQNFN 1620
QY 168 -----H-----E-----S-----EV-F----- 173
Db 1621 PIPNERLNAEVHHTDNIYERFANSHLTYRNEFQWVDSLKYTMGKGEVRFVSMKDLIDLID 1680
QY 174 -----I---C-----C-----C-----C-----C----- 175
Db 1681 GTLQAIVGCFYFFENTNDNSFPVFTIDQILNGDISQQLHAVALKYDSSGNFINGDATV 1740
QY 176 -----A-----A-----A-----A-----A----- 176
Db 1741 YDALGNIIILHISNVTKRLNGQSPSLTSKTVDSKTKKVENEDQKRASKNMLHWVFEE 1800
QY 177 -----W-----N-P-----V-----S-DLLASG----- 187
Db 1801 NFGWTIDIDNTTGFDFDLGLTSIQAVKLUNAISKSNYPNASSTCVFDPYSIDLL-SCYLSTLN 1859
QY 188 -----S-G-----D-S-----T-----A-R-----I-----W----- 196
Db 1860 DPQVETSTGDDIQKOLTEDHKPTRLAENPIGVMAAACRLPGVSSPSSELWELLKIGKN 1919
QY 197 -----N-L-----S-----S-----S-----S-----S----- 199
Db 1920 ASSRIPATVPTRNTLISGSKYGNPVEGCFITQDVTQFDPSPFFKISKSEAEILDPQORL 1979
QY 200 -----ENS-----TS-----GS-----GS----- 206
Db 1980 LLECQCELENGSVIETSNVGVFVGLMEKEXYQDMMESSILAMLGSMMAAIAGRVNVIFG 2039
QY 207 -----L-----VL-----R-----R-----T-----Q- 208
Db 2040 CYGPSVTIDTACSSSLVALEMAINALLDNRCISKVIVAGVNLILNEKGOGLRTNGKLSQH 2099
QY 209 -----L-----L-----L-----L-----L-----L----- 212
Db 2100 GMSLSDFRASGYGRSDGCVLMLLELAKFNHYMSTIQSVNVNHGGRSVSLTAPNGVAHK 2159
QY 213 -----H-----H-----H-----H-----H-----H----- 213
Db 2160 MLTTSVINGQPSLAIDYWEAHGTGTPLGDPIDBNTLSSILQNIIGSVKASLGHGEASAG 2219
QY 214 -C-----IR-----E-----G-----G-----G----- 219
Db 2220 TCGLLKLFMLTYQVPTLIHFHVLNKOINAGSIRLPIIGEDSELVSAGISSFGVSGTNA 2279
QY 220 -----Q-----DVP-----DVP-----DVP----- 223
Db 2280 AAIAFNDNNKLEPYPIHKYIYILPISAKNQISLDNLEKQILSVIPLTDVPICNIASALAN 2339
QY 224 -----SN-----KD-----KD-----KD----- 227
Db 2340 NRSHFTIRNALIVNSGIVNSKMEGKPHRVAKKDRVHVVKLDCSLDASLLQYDVINETYT 2399
QY 228 V-----TSL-----D-----W-----N-S-E----- 236
Db 2400 VASLKNQSFAMKFAIIRKFLTSLSLEYIEIVASDGBELLAVLLANGSLKWFENFKTMIELP 2459
QY 237 -CTL----A-----TG-----TG-----TG-----S 244
Db 2460 IGSLLTEFADHDLNSTSSSIKSYQTPESHNDLSPMELMKLIMKLYITGYOVDMWATVYS 2519
QY 245 -----Y-----D-----G-----F-----F-----F----- 248
Db 2520 PVEQFIALPNYQFNKOTLWFEERLEIVDHYLGTIDEBSEDTLILKNQISELRHHPFKG 2579
QY 249 -----A-R-I-----W-----T-----T----- 253
Db 2580 KPLDVGTWSEIAIBALKIRNEIPFSIQNLKLTBLITLTKPAWLETNVTRNEDDEGFNVSAY 2639
QY 254 -----K-----K-----K-----K-----K----- 254
Db 2640 IDQORLFSLNASSVEIQIEVPAVEVQIPDKVYVYLKCECPNAVIRHRNMVYVDSRAEQSP 2699
QY 255 -----DG-----DG-----DG-----DG-----DG----- 257

Db 2700 FRTANIVLEIGFAPTPSDMEILGLVPSVHYMVQVDDGALWQFQWISQDKREVLSNIY 2759
QY 258 -----LG-----LA-----
Db 2760 VLKDAKGLEIPTIRMHKKSTLLSQEASIVAAKTQMAVRHKVCLAVGDVIESGLDIDES 2819
QY 260 --ST--LG-----
Db 2820 QLSTGFSELGDSLATVDLLNRLNQYPPPEIELTTSDFPNPSIIDLSIMEOLLNEKGI 2879
QY 264 -----Q-----H-----
Db 2880 TEPSEPNTKSLRGRKLSIPAVRAQVLAQIEFVENVNSKQKEVQAEAPSSSECSNMLE 2939
QY 266 -----KG-----P-----
Db 2940 ESDATVDRTEIRRKVSLAVFDLATETLSAEDLOSQKGTGLGMDLSIVDFNRLNDKYFP 2999
QY 269 -----IF-----A-----271
Db 3000 DDEITASDIFDPTVDELSDHIVRKKSVPSPAASEIMKTMNGISTSVDAETHLENLS 3059
QY 272 -----LK--W--N--K--K--GNFI--LSA-----G-----285
Db 3060 QSFMLLENQNSINPTLKWINSQTIKLVKPSDGNFLFELNANGGQKEIKHFTGPNII 3119
QY 286 VD-K-----T-T-----I-----291
Db 3120 IDLKGHEGSTETLYMSLLNLVKSISKLEIQCRFGVSEQFGLGNSISRAFMTVAAEKNP 3179
QY 292 I--W-----D-----294
Db 3180 LISFAMVQNVQSVFVDSDSPTGNWLTGGLSGIGLEIGKFIANNAENVILISRRQPT 3239
QY 295 --A-----HTG-----298
Db 3240 AKALRDILSTELTHIGLAKTIVLKLIKNISAKLIVQSPKLSFTKYISDLPSKKVTF 3299
QY 299 -----E-----299
Db 3300 FYNLOSCKPSKHFHMLFEWLEVYLIKQMTVHTTAADINDKEKLIRELTKLVNGITGIH 3359
QY 300 -A-----K--Q--O--Q--P--P--PH-----307
Db 3360 SAGVLKSKIERQKESFNQVFTPKANGPHVLEEIEKHFNKYTIENFIMMSFTAAACNEG 3419
QY 308 -----S-----308
Db 3420 QLVGVSNAYLEYQVORRRQKSGCAIQWGNWIDTGMATDENVRKFLANLGLGQHNKD 3479
QY 309 -----A-----PAL-----DVM-----Q-----S 318
Db 3480 ALKYLACILTKPELIMVANIDWNVLKNRKLDPKOLINTGILPFEDFTGKNNESEFPPLS 3539
QY 319 N-----319
Db 3540 NGDFEKVMNFSVEDEEVLLEIKKVSSILMCSPTKLKNNKIMDMGLDSKLIIVBFLNF 3599
QY 320 -N-TF-----A-----S--C-----325
Db 3600 INSTFKISVNLSDAYNHPTLEKLAHI FEQMTIVDHPVNSKSEIEFKSTDFCPIFGINI 3659
QY 326 -----ST-----D-----328
Db 3660 FFDKNDFDRAKSTAVKLENGBQLPTAGYAVSVVGKSIIRDVVSKIEAAPQOIKLCOE 3719
QY 329 -----M-----C-----I-----331
Db 3720 SSKCVLMLTCQSQYPMNGQLVENYEI FRTTLOSCLKCDEYLGQDVLSLWEILFNTDHY 3779
QY 332 -----H--V--C-----334

Db 3780 KLLQLTKHQPIMFCFYATAQLWLSIGIVPDYVLGHSVGLVAGVLGIMSEDGLRLI 3839
QY 335 -----K-----LG-----Q-----D--R-----P--I--KT-----F- 345
Db 3840 VERGKAMENIAGLALLAQOREIADEVLRKFVKVSVATINSKQVVFAGTKSVLDAALAFV 3899
QY 346 --QG-----HTN--E-----VN-----A 354
Db 3900 KGCKQATVYNOQYPPHSLNIQETHLVLSRQCLADIKFSAGRTPLVSNVTGQIINFSEA 3959
QY 355 -I-----K--W--D-----PT-----360
Db 3960 YIVKHTVSAVKFVDCVETLQAKGVTWIDAGSAAVLATFVKRIIQPTLSKHRIVOTCKE 4019
QY 361 -----GNLL-A-----S--C-----S-----368
Db 4020 KESDVNDNLVQACLBLEQSGLPISWTTLYGCGRNADERLVEFPMTHNDIINKNDFELLEGH 4079
QY 369 -----DMT--D-----D-----372
Db 4080 QLNKGIWVAGAYQLFKIDQLVKLKAAGMELMLKNVKFLKPWYIEDNREYQIOWNSDWIE 4139
QY 373 -----LK-----374
Db 4140 LIVNSVIVCSLEVEPQNSVLKLETISENEKPEVHDFYETLFRNGLYQDSGFRRIESARR 4199
QY 375 -----I--W-----S--MK-----379
Db 4200 SDKRCFSQIKSSPFAWPLIDSAMHSITASVVRPRPCYFLPVAMGVTWMTKDTNSFTL 4259
QY 380 --Q-----D--N--C-----VH-----385
Db 4260 HAQTUITSETDKFIQNVNALLAGDTPICEVRNMTIVVLKTEPVHTRIPNSIETVETPPKS 4319
QY 386 -----D--L-----Q-----Q--HN--K-----392
Db 4320 EIEIVGDISLPYNOISENSENWOHLKTNTVKQLHNRSLKQDHARVALLDSDARYWDE 4379
QY 393 -----E-----I--YT-----I-----K 398
Db 4380 YGIRESEAKFIDPOORLLCSVAKLLDLSLITSLTNTGVFGCSANEFSHVYAYGK 4439
QY 399 -----WS-----P-----TG-----403
Db 4440 DPRAEWSGTSNSALAGRIAHWLKLGPPVTLDTACSSSFYALSACDALRTQCEYAI 4499
QY 404 -----P-----404
Db 4500 GTVNLVHMTTDLVQNAKMTVDDFCKAFDNDANGYKRYSEAVGCSMLLTSPNIDSVATIT 4559
QY 405 -----GT-----N-----NP-----NA-----411
Db 4560 NYATGHNGTSSSLFTPNGLSQLEVMOARATNPLEKILEIQTHCTGTGLGDPPIEINASKLV 4619
QY 412 -----N-----LM-----414
Db 4620 SSACKIGSVKSNIGHTEGSSGLVSLCSSLMSFRSKYRVAQLHLKCTNSIKTNKMICRFI 4679
QY 415 -----L-----A-----SA--S-----419
Db 4680 GEDADENNSILINNFQGTGNSCVLKPKNNAISEHFPVSSEVFPYILLSSHSKSLQKYVQ 4739
QY 420 -----F-----D--S-----T-----VR-----425
Db 4740 VLCEPISNSAKSLHDIMMSLFQKKIHVHRQFIIFNPKRIAVTSLDGFVEVRDERLEK 4799
QY 426 -----L-W-----D-VDR-----CI-----433
Db 4800 HPCSFLKEGVVHFDKXSFQFVDLPFPIVFNNTLHWALDSYRDEIDHESQMSFKNIFY 4859
QY 434 -----CI-----H-----436
Db 4860 EKVLLETMPQNPISKVVICIGRLDLIPKIEDEVSSFANGIIVFPHIPNISIFEYLK 4919

QY 437 --TL--TKH--Q--E--P-----V----- 445
Db 4920 VMSLISRNQNVFIICCFENGTSHTWTGTLRLSLASEKMFYKFVSDIKVDALKLFENHE 4979
QY 446 -----YS-----V-AF-----S-----P----- 452
Db 4980 YMFEAIFYKRGYRVERLRKVNPTFKPAQYKCLISGGTGGIGSAIINELKPKSSVIT 5039
QY 453 -----DG-----R-----Y-----LASG-----SFD----- 463
Db 5040 RKNIASEDGKTFUSSDITRLDISHKFNYPHLA-GIVNNSLHENVKRDSLDENVSKLOG 5098
QY 464 -----RC-----||-----VH----- 467
Db 5099 AKULMKCCDTSHFVSSSIANVLGSYGQSYAFNSGLVTSFLETSSTKSTIIHWGPWKD 5158
QY 468 -----I-----W-----N-----TQ-----T-----G-----ALV-----H 478
Db 5159 VGMLAQPERRERIVKQIESNGWKLPPNQDAISVFYTFQMETHEQIIIVFDGDFDTIVARQPH 5218
QY 479 -----S-----Y-----TG----- 478
Db 5219 LOKLSEVVEKTKVKEIKKKSLNFEIPEIIVGIIITDSSKLNIPFMDLGIDSLCMLNLR 5278
QY 479 -----S-----Y-----Y-----Y-----Y-----R----- 481
Db 5279 YSLNKNFDLELTVSEMFENATYKLTQYVETLRLKAKHSLSLSDHRYSSQVSNKEDTRVA 5338
QY 482 -----G-----TG----- 484
Db 5339 VIGWAEFGSSNIHEYWENLMDGICSTGNKYLKPNFGDNKFNFLTDEDARVLDPOV 5398
QY 485 -----GIF--E-----V----- 489
Db 5399 RXPQHAYLALENSGVYKQKHELRCGVFAGAEPSDYGRADDDHDDAMKLFVMMNMSYLAS 5458
QY 490 -----C-----W-----NA-----AG--D-----KV--CA----- 500
Db 5459 YASYCLDLKGEAVSVYSACSTALVANAVAKSIQSGSMYALVGAASIAEVSGALSGFDD 5518
QY 501 -----SAS-----D-----GS--V--C--VL-----DL----- 512
Db 5519 QKKTMPKSGVCRPFDPKDEGIVRGSGVGVFLKRYSQLLDNDNVHVFVKDFAINNDGH 5578
QY 513 -R-----K 514
Db 5579 SRASFAPNPAGQLK 5593

RESULT 9
T42737
gp330 protein precursor - rat
N:Alternate names: megalin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
R:Accession: T42737
R:Saico, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A:Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of the gp330 family
A:Reference number: A58173; MUID:95024033; PMID:7937880
A:Accession: T42737
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4660 <SAI>
A:Cross-references: UNIPROT:P98158; EMBL:L34049; NID:g561852; PID:g561853; PIDN:AAA51369
A:Experimental source: strain Sprague-Dawley; kidney
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
F:1-25/Domain: signal sequence #status predicted <SIG>
F:46-4660/Product: gp330 protein #status predicted <MAT>

Query Match 70.4%; Score 2574.9; DB 2; length 4660;
Best Local Similarity 9.8%; Pred. No. 2e-52;

Matches 426; Conservative 66; Mismatches 14; Indels 3841; Gaps 368;
QY 1 M-S--I--S--SDE-----V----- 8
Db 113 WTCNSQCIPSEYRCDHVSDCPDGSDERNCHYPTCDQLTCANGACACNTSORDCQKVDGRD 172
QY 9 ---N---F---L-VY---R---Y--- 15
Db 173 SSDEANTTLCQKEFGCGECILRAYVCDHDNDCEDNSDRCNRYDTCGGHQTCSNG 232
QY 16 --L-----Q-----ES----- 19
Db 233 QCINQNVWCGDDDCQDSGDEDCESNQSHRCYPREWACPGSGRCISIDKVCDCGVDPDCP 292
QY 20 ---G-F-----S----- 22
Db 293 EGDENNVTSGRTCGMGVCSVLNCEYQCHQTFPGGECFCPPGHIINSNDSRTCIDFDDCQ 352
QY 23 ---H---S-AFT---F-----G--- 29
Db 353 IWGICDQKCNQRGRHOCLECEGYILRGHCHCKSSDSFSAASVIFSNGRDLLVGDHLGRN 412
QY 30 -I--KS-----H-----I---SQ---S---N-- 38
Db 413 FRILAESKNRGMVGVDFHYQKRVFWTDPMQEKVFSTDIINGLNTQEIILNVSVDTPENLA 472
QY 39 --IN-----G-AL--V----- 44
Db 473 VDWINKLYLVETKNRIVDVNLEGNQVRVLTITENLGHPRGIALDPTVGYLFFSDMGSL 532
QY 45 --P-----PA---ALIS-----I-----IQ-K- 55
Db 533 GQPKVERAFMDGSRKDLVTKVGHGAGITLDLVSRYVWDSRYDIETVTVDGIOKRT 592
QY 56 ---G--L-----Q-----Y--- 59
Db 593 VARGSLVPHFPFGISLFEHVFFTDWTQAMVMAKSKFTETNPQVYHQSLLRPHGVTVYHA 652
QY 60 ---V-----E---AE--V81-N--- 67
Db 653 LRQPNATPCGNNGGCAQCVLSHRTDNGGLGYRCKEFGFELDDDEHRCVAVKQFLLF 712
QY 68 -----E-----DGT 71
Db 713 SKTAVRGIPFTLSQEDVMVPVTSPPSFVGVGIDFDAQHSTVFSYSLSKDIIYKQKIDGT 772
QY 72 -----L-FD----- 74
Db 773 GKEVITANRLSEVCLTFDWISRLNLYWTDGGLKSVTLRLADKSRRIISNLNPRISV 832
QY 75 ---G---RP--I---E-S-L---SL---I-----DA--- 86
Db 833 HPTAGYMFUSDWFRPAKIMRAWSDGSHLMPIVNTSLGWPNGLAIDWSASRLYVWDAPFDK 892
QY 87 -----V--M--P-----D-V-----VQ-----T--R- 95
Db 893 IEHSTLDGLDRKRLGHVDQTHPPGLTVFKDNVFTDWRGAIIRVKSDDGDMTVIRRG 952
QY 96 -----Q-----Q---A-----Y-----RKL-- 103
Db 953 ISSVMHVKAYDADLQTSNYCSQTTHANGDCSHFCFPVPNFQVRCVCPYGMKLRQDMTC 1012
QY 104 ---A---QQ-----Q-----A-A 109
Db 1013 EGDPAEPPTQCGSLSPFCNNKGVSPFRCDGVDDCHDNDSEHOCGVFNNTCSPSFA 1072
QY 110 -----A----- 110
Db 1073 CVRGQCIPQWHCDRQNDCLDSDEQNCPTATSTCTPSTFTCDNHVCIPKDWCDTD 1132
QY 111 -----A-----A-----A-----AAA 116
Db 1133 NDCSDGSDKNCQASGTCQPTQRCFDHRCISPLVYVCDGDKDCADGSDGACVNLNCTSAQ 1192

QY 117 ---A--AS-----Q-Q----- 121
Db 1193 FKADGSSCINSRYRCGVDCRNSDBAGCPTRPPGKCHPDEFQCGDGTCPINTWECD 1252
QY 122 ---GS---A-K-----NG-----E----- 128
Db 1253 GHPDCIHGSDHETGCVKTPCSPTFLCDNGNCIYKAMICDGDNDCRDMSDEKDCPTOPFH 1312
QY 129 ---N-TA-----NG-E-E-----NG-AH----- 139
Db 1313 CPSTQWQCGYSTCINLSALCDGVDFCPNGTDESPLCNQDSCHFNCGCTHQCWQGPFGA 1372
QY 140 ---T-----I-----A----- 142
Db 1373 TLCPLGYQLANDTKCEDINECDIPGFCQHCVMNRGSRFCACDPBYTLESRGRTCKVT 1432
QY 143 ---N-----N-HT-----DMW----- 149
Db 1433 GSENPPLVVASROKIIVDNITAHNLYSLVDVSVFVALDFDSVTCGRVFSWDLLOQKWTW 1492
QY 150 ---E---VD-G---D---VE-----I----- 157
Db 1493 SVFQNGTKRVRVHDSGLSVTEMAVDWIGNLYWTDYALETIEVSKIDGSHRTVLISKNV 1552
QY 158 ---P---N-----KA-----VV-----L-----R- 166
Db 1553 TKRGLALDPRMGDNVWFSDWGHHPRIERASMDGTWRTVIVQEKIYWPCLSIDYPNRL 1612
QY 167 ---GH-E---SE-V-----F----- 173
Db 1613 IYFMDAYLDVIEFCDYDGHNRROVIAISDLVHLHPHALTLFEDFYWTDRTQVMOQANKW 1672
QY 174 ---I-----CA-----W 177
Db 1673 HGGNQSVVMYVHQPLGITAHPSPRPPRNPSCASCSHLCLLSAQAPRHYSCACPSGW 1732
QY 178 N-----PVS-----D----- 182
Db 1733 NLSDDSVNVCYRQDFLMSVRDNIIFGISLDPEVKSNDAMVPISGIGHQYDVFDFSEQF 1792
QY 183 -L-----LA-S-G-S-D-S-----T-A----- 193
Db 1793 IYVVENPGEIHRVKTGDSNRTVEAPLSGLSLGLALDWVSRNIYYTTPASRSIEVLTK 1852
QY 194 ---R-----I-----W----- 196
Db 1853 GDRYKGLTIANDGTPLVGFPVGVIAVDPAARGKLYMSDHDGTSVPKAKIASANMDGTSK 1912
QY 197 ---NL-----S-E-N-ST-----S-----GS 206
Db 1913 ILFTGNLQHLLEVTLDIQEQKLYWAVTSRGVIERGNVDGTERMLVHHLAHPWGLVYGS 1972
QY 207 ---T-QL-----VLR-----H-----C-- 214
Db 1973 FLYVSDQYEVIERVDKSGNKKVLRDNVPLRGLRVYHRRNAADSSNGCSNPNACQQ 2032
QY 215 I-----R-----E-G----- 218
Db 2033 ICLFVPGMFSCACAGFKLSPDGRSCSPYNSFMVMSLPAVRGFSLELSDHSEAMVPVA 2092
QY 219 GQ-----DV-----P-SN-----K-- 226
Db 2093 GQGNVLHADVDVANGFIYCDPSSSVRSNGIRRIKPDGNSFTNVVYIGANGIRGVA 2152
QY 227 -D-----V-TS-----L-----D-----W----- 233
Db 2153 LDWAAGNLFTNAPVYETLLEVLRIINTYRVLVKVSDMPRHIIYDPKHYLFWADYQ 2212
QY 234 -----N-----SEG-T---LLA---TGSY-----DG----- 247
Db 2213 KP KIERSFLDCTNRVLVSEGIYTPRGL-AMWDHTG-YIYVWDDSLDLIARIHLDGSESQ 2270

QY 248 ---P-AR-----IWTKD-----GN----- 257
Db 2271 VRYGSRYPYGITVFGESIHW-DRNLKKVFAQSQGNTDPPVVIRDKINLLRDVTI 2329
QY 258 ---L-A-----S-----TLG-----Q-- 264
Db 2330 FDEHAQPLSPAELNNPCLQSGGSHFCFALPELTPRCGCAFGTLGNDGKSCATSQED 2389
QY 265 -----H-----K----- 266
Db 2390 FLIYSLNLSRLSHLPRDHSPLFPQVISVAGTAIALDYDRNNRIFFTOKLSLRGQISY 2449
QY 267 ---G-P-----I-F----- 270
Db 2450 VSLYSSSPTVLLSNIGVTGDIAFWINRIYYSFQNTNSMAEDSGSNRAVIARVSK 2509
QY 271 -A-----LK-W-N-K-K---GNF---IL-S---AG-----V 286
Db 2510 PRAVLDPCKGYMYMTDWTGNAKIERATLGGNFRVPIVNTSLVMPNGLALDLETLLYWA 2569
QY 287 D---K-----T-----TI---I-W-----D--A 295
Db 2570 DASLOKIERSTLTGTNRVSVTAHSGFLTVYGOIYITDLYTRKIYRANKYDGSDLVA 2629
QY 296 -----P-----A-----L----- 312
Db 2630 MTRLPQPSGISVTVKTOQQCSNPDOFNGCCHICAPGNGAECQCPHEGNWYLAND 2689
QY 301 -K-----Q-QF-----P-----PH-S-A--- 309
Db 2690 NKYCVVDITGRCNQLQFTCLNGHCINQDWKCDNDGCGSDELPTVCFAFHTCRSTAFTC 2749
QY 310 -----P-----A-----L----- 312
Db 2750 GNGRCVPYHYRCDYNDGNSDEAGLFRNCNSTTFTCSNGRCIPLSYVVCINNCHD 2809
QY 313 -----D-----V-----D----- 315
Db 2810 NDTSDKNCPPHTCPDFTKCTQTNICVPRAFLCDGNDGDSDENPIYCASHTCSNE 2869
QY 316 -----W-----Q-----S----- 318
Db 2870 FOCLSPQRCPISYWFCDGADGADGDEPDTCGHSVNTCRASQFQCDNGRCISGNWVCDG 2929
QY 319 -N-----N-T-F----- 322
Db 2930 DNDGGMDESDORHCELQNCSSQTCTVNSRPPNRRRCIPQYVWCGDADCDALDELQ 2989
QY 323 -----A-SC----- 325
Db 2990 CTMRTCSAGEFSCANGRCVRSQSPRCDRRNDGSDYSDRGCSYPPCHANQFTCQNGRCIPR 3049
QY 326 -----S-----T-----D-MCI-----HV-C-----K 335
Db 3050 FFVCDDBDNGDGSDBEQHCHTPTCPLHQFRCNDGHCIEGMVRCNVHVDSDSNDSEK 3109
QY 336 -----L-G-----O-- 338
Db 3110 GCGINECLDSISRDHNCTDTITSYCCLPGYKLMDSKRCVCDIDECKESPOLCSQKC 3169
QY 339 -----D-R-----P-I----- 342
Db 3170 ENVVGSYICKCAPYIREPDGKSCRONIEPYLIFSRYRINLTDDGSSYSLLIQLGLG 3229
QY 343 -----KT----- 344
Db 3230 NVVALDFDRVEKRLYWADEAKQIIBRMFLNKTNRETIINHRLRAESLAVDWVSRKLYWL 3289
QY 345 -----F-----O-GH----- 348
Db 3290 DAILDCLFVSDLEGRHRKMAIOACHVDANNTFCFHEPRGIVLHPQRGHVWADGWVHAYIG 3349
QY 349 -----TNE-V-----NAI-----KW-----D----- 358

Db 3350 RIGMDGTNKSIIISTKIEWPNAITIDYTNLLYWAHAHLGYIEFSDLEGGHRRHTVYDGLS 3409
QY 359 --P--T--GN--L--LAS- 366
Db 3410 PHFPALTIPEDTVFTWTDWNRTRTEKGNKYDGSRRVVLVNTTHKPPDIHVHPYRQPMIN 3469
QY 367 ---CS---DD---M---T---L---K--- 374
Db 3470 PCGTNNGGCSHLCLIKAGRGFTACPDPTQVQLRDRTLCPMPCSSSTQFLCGNNEKCP 3529
QY 375 IW---S---MK---Q---D---NC---V 384
Db 3530 IHWKCDGQKDCSGSDPDLCPHRCFLGFCQCRDGNCTSPQALCNARQCADGSDSDRV 3589
QY 385 ---H---D---L---Q---Q---H---N--- 391
Db 3590 LCEHHRCESENEWQCANRCIPQSWQDCSVNDCLDNDEDTSHCATCRPGQKCNNGRC 3649
QY 392 ---K---E---I---Y---T---IK-W--- 399
Db 3650 IPQSWKCDVNDGCDYSDEPIDECTTAAYNCNDNHTEFSCKTNVRCIPQWAVCGNPGDDCD 3709
QY 400 ---S---PTG---P---GT---N---N--- 408
Db 3710 NSDEQCESVPCHPSGDFRCANHHCIPLRWKCDGTDGCGNDSDEENCVPRECSSEFRCA 3769
QY 409 ---P---N---AN---LM---LA---SAS 419
Db 3770 DQOCIPSRVWQDENDCGNDSRDERCEMTKCHPEHFQCTSGHCVCPKALACDGRADCLDAS 3829
QY 420 FD-S---TVR---LW---D---VDRG---I---C---I--- 435
Db 3830 -DESACPT-RFPNGTYCPAAMFECKNHVCIOQPMICDGENDCVD-GSDEIHLCFNIPCE 3886
QY 436 ---HTL---T---K---H---QEPV---Y---S--- 447
Db 3887 SPORFRCDNSRCVYGHQLCVNGVDDCGSDGSEKEHCKRPHKCTDTEYKCSNGNCISQH 3946
QY 448 ---V---A---A---P---S---P--- 452
Db 3947 YVCDNVNDCGLSDSETGNCNLGNDRNRTCAENICEONCTQLSSGGFCISCRPGFKPSTSDKNS 4006
QY 453 --D---G---RYLAS-GS--F--D--K---C---C---V 466
Db 4007 CQDINECEBFGICQSCR--NSKGSYECFCVDGFKSMSTHYGERCAADGSPPLLLPENV 4064
QY 467 ---HI---WN---T---Q---T---GA--- 475
Db 4065 RIRKNTSSEKFESEYEEBEEHIQTIDYDWDPEHIGLSVVVYTVLAQSGQFGAIKRAYIPN 4124
QY 476 ---L---V---H---S---YR--- 481
Db 4125 FESGNNPIREVDLGLKYLMPDGLAVDWGRHYWSDAKSORIEVATLDGRYRKWLIT 4184
QY 482 ---G---T---G--- 484
Db 4185 QLDQPAIAVNPGLMFWTDQKQPKIESAWNGEHSVLYSENIGWPNGLSIDYLND 4244
QY 485 ---G---IFE---V---C-W---NAAG- 495
Db 4245 RVYWSKEDVIEAIKYDGTDRRLIINEAMKPPSLDIFEDKLYWAKEKEGVWRQNKFGK 4304
QY 496 ---DKV---GA-S-A---SD---G 505
Db 4305 ENKEKVLVYNPMLTVQRIHQRLAYNSVSNPCKQVCSHLCLLRPGGYSACPCQSGDFVTG 4364
QY 506 S-V-C---V---LD--- 511
Db 4365 STVQCDAASELPTMPPPCRMHGGNCYFDENELPKCKSSGSGYCEVGLSRGIPPGT 4424
QY 512 -----LRK 514

Db 4425 TMAVLLTFVIVIVGALVLVGLFHYRK 4451
RESULT 10
A88852
protein unc-22 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: A88852
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: A88852
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <STO>
A:Cross-references: UNIPROT:Q23550; GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN000
C:Gene: unc-22
A:Map position: 4
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology; I
Query Match 70.3%; Score 2572.2; DB 2; Length 6831;
Best Local Similarity 8.0%; Pred. No. 5.1e-52;
Matches 472; Conservative 29; Mismatches 10; Indels 5358; Gaps 394;
QY 1 M-----S-----I-----S-S-----D----- 6
Db 784 MTEAERRQSLFPQKVKEMWDIPLPEKTQQQVDKICEWKCTYSRPNKIRWYKDKKEIFS 843
QY 7 -----EV-----N-----FL----- 11
Db 844 GGLKVKIVIEKNVCTLIINNPEVDDGTGKYTCBANGVTHAQLTVLEPPMKYSFLNPLNT 903
QY 12 --VYR-----Y-----L----- 16
Db 904 QBIYRTQAVLTKVNTTPRALVWYRGSKAIQEGDPRFIEKDAVGRCTLTIKEVEEDQ 963
QY 17 -----Q-----ES----- 19
Db 964 AEWTARITQDVFQKQVYVEEPRHTFVVPKMSQKNESDLATLETVDNDKDAEVMWHDG 1023
QY 20 -----G-F--S-----H-----S-----A-F----- 26
Db 1024 KRIDIGVKFKVSSNRKRLLIINGARIEDHGEYKCTTKDDRTMAQLIVDAKNKFIVALK 1083
QY 27 -----T---F-G--I---K---S---H----- 33
Db 1084 DTEVIEKDDVLMCQTKDTKTFGIWFRNCKQISSMPGGKFETQSRNGTHTLKIGKIEMNE 1143
QY 34 -----ISO-----S-----N----- 38
Db 1144 ADVYEIDQAGLRGSCNVTVLEAKRPILNWKPKIEAKAGEPCVVKVQFIKTRRGDPK 1203
QY 39 --I--NG----- 43
Db 1204 AQLKNGKPIDBEMRKLVEIIKDDVAEIVFKNPQLADTGKWALELGNAGTALAPFELP 1263
QY 44 V-----PP-----AA--LI--S-----II--Q-----KG 56
Db 1264 VKDKPKPKGPLETKNVTAEGLDLVWGTPDDEGAPVKAYIIEMQGRSGNWKVGETKG 1323
QY 57 -----L---QY---VEA-----E-V-----SI-----N----- 67
Db 1324 TDFKVKDLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPYGVGPKPKNMDAIDVDKHC 1383
QY 68 -----EDG-----T-----L----- 72
Db 1384 TLAWPEPDEGGAPITGYIIERREKSEKDWQGVQTKPCCCLTDKVVVEDKEYLYRVA 1443

QY 73 -----F-----D----- 74
Db 1444 VNKAGPCDGHKPIKWKAKKASPEFTGGGIGKDLRLKVGETIKYDVPISGEPLPECLWV 1503
QY 75 ---GRP---IE---S---L---S--- 82
Db 1504 VNGKPLKAVGRVKMSSRGKHKIMKENAVRADSGKFTITLKNSSGSCDSTATVTVGRPT 1563
QY 83 ---L-I---D-AVM---PD---VV-Q---T- 94
Db 1564 PPKGPLDIADVCADGATLSMNPDDGGDPLTGYIVEAQDMNDKGYIEVGKVDNPTTIL 1623
QY 95 -----R-----Q-O----- 97
Db 1624 KVLGRNKGNYKFRVAVNNEGESEPLSADQYTOIKDPWDEPGKGRPEITDPPADRIDI 1683
QY 98 -----A---Y---RD----- 101
Db 1684 AWEPPHKDGGAPIEIEYIEVRDPTDKEWKEKVRVPTDNASISGLKEGKEYQFRVAVNKA 1743
QY 102 -----K-LA-----Q 105
Db 1744 GPGQSPSEKQLAKPKFIPAWLKHDNLKLSITVAGATVRWEVKIGGPIPEVKWFKGNQ 1803
QY 106 Q-O-----A-----A-A----- 110
Db 1804 QLENGIQLTIDTRKNEHTILCIPAMRSDVGEYRLTVKNSHGADDEKXANLTVLDRPSKN 1863
QY 111 -----A-----A-----A----- 113
Db 1864 GPLEVSDVFEDNLMSWKPPDDGGGPIEYVEVEKLDATGATGRWVPCAKVKDKTAHIDGLK 1923
QY 114 -----A-----A-----A----- 118
Db 1924 KGQTYQFRVAVNKEGASDALSTDKTKAXNPYDEPGKTGTGTPVDVMDADRVSLEWEPK 1983
QY 119 S-----Q-----Q-G-S-----A-----K-N-----GE 128
Db 1984 SDGAPITQYIEKKGRGRDQECGVSGDQTNABEILGLKEGEYQFRVAVNKAAGPGE 2043
QY 129 -----N-----T-AN-----GE-----E-N----- 136
Db 2044 ASDPSRKVAKPRLKPIWDREAMKTTIKVGNDEVPDVPVGEPPPKKEWIFNEKFPVDD 2103
QY 137 -----GA-H-----TI-A-N----- 143
Db 2104 QKRIESEDYKTRFVLRGATRKAGLYTLTATNAGSGDKHSVEVILGKPSPLGLEVS 2163
QY 144 N-----H-----TDM-----ME----- 150
Db 2164 NVYEDRADLEWKVPEDDGGAPIDHYEIEKMDLATGRWVPCGRSETTKTTPNLPQGEYK 2223
QY 151 V-----D-----G-----D-----V 155
Db 2224 FRVAVNKEGESDPLTTNTAILAKNPYEVPGKVDKPELVMDKDHVDLAWNAPDDGGAPI 2283
QY 156 E--I-----P-----P-----NKA-----V 163
Db 2284 EAFVIEKKDKNRWEALVVPQGOKTATVBNLKEGEYQFRISARNKAGTGDSPSDRV 2343
QY 164 V-----L-----R-----G-----H-----ES----- 170
Db 2344 VAKPRNLAPRIHREDLSDTTVKVGATLKFIHVHIDGEPAPDVTWSFNGKIGESKAQIENE 2403
QY 171 -----F-----I-----C-----EV----- 172
Db 2404 PYISFALPKALRKQSGKYTITATNINGDSVTINIKVSKTPKPKPIEVTDFEDRAT 2463
QY 173 -----F-----I-----C----- 175
Db 2464 LDWKPPDDGGEPIEFIEKMNKDGIVWPCGRSGDTHFTVDSLNGKDHKFRVAVNS 2523
QY 176 -----G-----Q-----HK-G-----P-----I----- 269

Db 2524 EGSPDPLETETDILAKNPPDRPDRPGRPEPTDWDSDHVDLKWDPPLSDGGAPEEYQIEK 2583
QY 177 -----W----- 177
Db 2584 RTKYGRWEPAITVPGQQTATVPDLTPNEEYEVVAVNKGSPSPSDASKAVIAKPRNL 2643
QY 178 -----N-----PVS-----D-----L----- 183
Db 2644 KPHIDRALKNLTKIKAGQISFDVPVSGBPAPTVTWHPDNREIRNGRGRVCLDNPEYQSK 2703
QY 184 L-----A-----S-G-----S-GD----- 190
Db 2704 LVVKQMERGDSGTFTIKAVNNGEDEATVKINVIDKPTSPNGPLDVSVDHGHVTLNWRA 2763
QY 191 -----S-----TA-----R----- 194
Db 2764 PDDGGIPIENYVIEKYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVAVNAEAGE 2823
QY 195 -----I----- 195
Db 2824 SDPLETFTGTLAKDPDFPKTNAPBITDWDKDHVDLEWKPPANDGGAPIEYVEMKOE 2883
QY 196 -----W-----NL-----S-E----- 200
Db 2884 FSPFNDVAVHPAGQTNATVGNLKEGKYEFRIRAKNKGAGLGDPSDSASAVAKARNVPPV 2943
QY 201 ---NS-----T-S-----GST----- 207
Db 2944 IDRNSTQIEIKVQAGQFSLNIPVSGEPTPTITWTFEGTPVESDDRMKLNNEDGKTFHVK 3003
QY 208 Q-L-----VL-----R----- 212
Db 3004 RALRSTGTGVIKAENENGTDTAEVKTVLDHPSSPRGPLDVTNIVKDGCDLAWKSPEDD 3063
QY 213 -----H-----C----- 214
Db 3064 GGAEISHYVIEKQDAATGRWTAGCESKDTNFHVDLDTQGEYKFRVAVNRHGDSDPLEA 3123
QY 215 -----I-----R-----EG----- 218
Db 3124 REAIIAKDPFDRADKPGTPEIVDMKDHADLKWTTPPADDDGAPIEGYLVEMRTPSGDWVP 3183
QY 219 ----- 218
Db 3184 AVTVGAGELTATVDGLKPGQTYQFRVAVNKAAGESTPSPSRMTWAKPRHAPKINRDMF 3243
QY 219 -----GQ-----DV-----P----- 223
Db 3244 VAQRVKGATLNFVDNVEGEPAPKIEWFLNGSPLSSGGNTHIDNNTDNNKLTTKSTARA 3303
QY 224 -----S-----N-----KQVT-S-----LD----- 232
Db 3304 DSGKYKIVATNESGKDEHVDVNILDIPGAPGELHAKDITKESVVLKWDPELDDGSGPI 3363
QY 233 -----W-----N-SE-----GT-L----- 239
Db 3364 TNYVVEKQSDGGRWVPCGETSOTSLKVNKLSGHEYKFRVAVNRQGTSAPLTSDHAIWA 3423
QY 240 -----L-----A-----T-----G-----SY-----DG 247
Db 3424 KNPFDEPDAPTDTVTPVDWKDHVDLEWKPPANDGGAPIDAYIVEKDKFGDWVECARVDG 3483
QY 248 -----F----- 251
Db 3484 KTTKATADNLTPGTYQFRVAVNKAAGPKSPDPTGNVVVAKPRRMAPKLNLAGLDRK 3543
QY 252 -----W-----TKD-G-----NL-----A-S-----T-L----- 262
Db 3544 AGTPIKLDIAPGEPAPVAKWKANDATIDTGARADVNTPTSSAHIHFAVRGDTGVYKI 3603
QY 263 -----G-----Q-----HK-G-----P-----I----- 269

Db 3604 I V E N E H G K D T A Q C N V T V L D V P G T P E G L K I D E I H K B G C T L N M K P P T D N G G T D V L H V I V E K 3663
QY 270 -----F--A-----L-----272
Db 3664 M D T S R G T W Q E V G T F P D C T A K N K L V P G K Y A F R V A V N I Q G S K P L E A E S P I I A K N Q P D V 3723
QY 273 -----K W N-----K K-----G-----278
Db 3724 P D P V D K P E V T D M K D I D I K W N P T A N N G G A P V T G Y I V E K K E G S A I W T E A G K T P G T P T S A 3783
QY 279 -N-----F-----I L--S-----A G-----285
Db 3784 D N L K P G Y E F R V I A V N A A G S D P S D P T D P Q I T K A R Y L K P K I L T A S R K I K I K A G F T H N L E 3843
QY 286 -----V D -K--T T-----290
Db 3844 V D F I G A P D T A T W T V G D S G A A L A P E L L V D A K S T T S I F P P S A K R A D S G N Y K L K V K N E L G E 3903
QY 291 -I--I-----W-----D-----294
Db 3904 D E A I P E V I Q D R P S A P E G P L E V S D V T K D S C V L N M K P P K D G G A E I S N Y V V E K R D T K T N T W 3963
QY 295 -A-----H-----T--G-----E--A K O O F-----304
Db 3964 V P S A F V T G T S I T V P K L T E G H E Y F R V M A E N T F G R S D S L N T D E P V L A K D P F G T P G K P G R P 4023
QY 305 -----P-----P-----F-----306
Db 4024 E I V D T D N D H I D I K W D P P R N G S P V D H Y D I E R K D A K T G R W I K N T S P V O G T A F S D T R V O K 4083
QY 307 -H-----S A-----P--A-----L D-----V-----314
Db 4084 G H T Y E V R V A V N K A G P Q S D S A A T A K P M H E A P K F D L D L D G K F R V K A G E L V I T I P F 4143
QY 315 -----D--W-----Q-----S-----318
Db 4144 T A S P Q D I S W T K E G K P L A G V E T T D S Q T K L V I P S T R R S D S G P V K I K A V N P Y G E A E A N I K I 4203
QY 319 -----N-----N-----320
Db 4204 T V I D K G A P E N I T Y P A V S R H C T L N W D A P K D G G A I A G Y K I E Y E V G S Q I W D K V P L I S 4263
QY 321 -----T-----F-----A--S--C-----S T-----327
Db 4264 G T A Y T V R G L E H G Q O Y F R I A E N A V G L S D Y C Q G V P V I K D P P D P P C A P S T P E I T G Y D T N Q 4323
QY 328 -----D-----D-----M-----C I-----331
Db 4324 V S L A M P P R D D G S P I L G Y V V E R F E K R G G D W A P V K M P M V K G T E C I V P G L H E N E T Y Q P R V 4383
QY 332 -----H-----V--C-----334
Db 4384 R A V N A A G H G E P S G S E P V T C R P Y V E K P G A D A P R V G K I T K N S A E L T W N R P L R D G G A P I D G 4443
QY 335 -----K L G-----337
Db 4444 Y I V E K K L G D N W T R C N D K P V R D T A F E V K N L G E K E Y E F R V I A V N S A G E G S K P S D L V L 4503
QY 338 -Q-----D-----R-----P I-----342
Db 4504 I B E Q G R P I F D I N N L K D I T V R A G E T I Q I R I P Y A G N P K P I I D L F N G S P I F E N E R T W D V 4563
QY 343 -----K--T-----F-----Q-----346
Db 4564 N P G E I V I T T G S K R S D A G P Y K I S A T N Y K G O T C K L A N V F L D A P G K T G P I R A T D I Q A D A M 4623
QY 347 -----G H-----T-----N--E--V--N--353
Db 4624 T L S W R P P K O N G D A I T N Y V V E K T P G D W T V G H P V G T I L R V N L A N T P Y E F R V A E N Q 4683
QY 354 -----A I--K-----W-----D-----P-----359
Db 4684 Y G V G P L E T D D A I V A K N P F D T P C A G P Q E A V E T S E A I T L Q W T R P T S D G A P I Q G Y V I E K 4743

QY 360 -----T--G N-----362
Db 4744 R E V G S T E W T K A A F G N I L D T K H R V T G L T P K K T Y E F R V A A Y N A A G Q G E Y S V N S V P I T A D N A P 4803
QY 363 -----L-----363
Db 4804 T R P K I N M G L T R D I L A Y A G E R A K I L V P F A A S P A P K V T F S K G E N K I S P T D P R V K V E Y S D F L 4863
QY 364 -----L-----A S-----C-----367
Db 4864 A T L T I E K S E L T D G G L Y F V E L E N S Q G S D S A S I R L K V V D K P A S P O H I R V E D I A P D C C T L Y W M 4923
QY 368 -S D-----D M-----T-----372
Db 4924 P F S S D G S P I T N Y I V E K L D R H S D K W E K V S S P V R N L N T V G G L I K D N R Y R F R V R A E T Q Y 4983
QY 373 -L-----K-----I--Q-----W-----S M K-----Q--380
Db 4984 G V S E P C E L A D V V V A K Y Q F E V P N Q P E A P T V R D K D S T W A E L E W D P P R D G G S -K I I G Y O V Q Y R 5042
QY 381 D-----N-----C V H D-----L--Q-----389
Db 5043 D T S S G R W I N A K M D L S E Q C -H A R V T G L R Q N G E F E F R I I A K N A A G F S K P S P P S E R C Q L K S R F 5101
QY 390 -----H-----N-----K-----E I--Y--T I K W-----S-----400
Db 5102 G P P G P I H V G A K S I G R N H C T I T W A P L E D G G S K I T G Y N V E I R E Y G S T L -W T V A S D Y N V R E 5160
QY 401 P--T-----G P-----G-----T--N 407
Db 5161 P E F T V D K L R E F N D Y E F R V V A I N A A G K I P S L P S G P I K I Q E S G S R P Q I V V K P E D T A Q P Y N 5220
QY 408 -----N-----P-----409
Db 5221 R R A V F T C E A V G P E P T A R W L R N G R E L P E S S R Y R F E A S D G Y K F T I K E V M D I D A G E Y T V E V 5280
QY 410 -N-----A N L-----413
Db 5281 S N P Y G S D T A T A N L V Q A P P V I E K D V P N T I L P S G D L V R L K I Y P S G T A P P R H S L V N R E E I D 5340
QY 414 M-----L-----A--S--A-----S-----F-----420
Db 5341 M D H P T I R I V E F D D H I L I T I P A L S V R E A G R Y E V T V S N D S G E A T F G M L N V T G L P E A P Q G P L 5400
QY 421 -----D--S-----T V-----424
Db 5401 H I S N I G P S T A L S W R P P V T D G S K I T S Y V V E K R D L S K D E W T V T S N V K D M N Y I V T G L F E N 5460
QY 425 -----424
Db 5461 H B Y E F R V S A Q N E N G I C A P L V S E H P I I A R L P F D P T S P L N L E I V O V G D Y V T L S W O R P L S D 5520
QY 425 -R L-----W-----427
Db 5521 G G R L R G Y I V E K Q E B E H D E W F R C N Q N P S P P N V N V P N L I D G R K Y R Y R F A V N D A G L S D L A 5580
QY 428 -D-----V-----D-----R-----431
Db 5581 E L D Q T L F Q A S G S G E G P K I V S P L S D L N E E V G R C V T F E C I S G S P R P Y R M F K G C K E L V D T S 5640
QY 432 -G-----I-----C-----H--436
Db 5641 K Y T L I N K G D K V L I I N D L T S D D A D E Y T C R A T N S S G T R S T R A N L R I K T K P R V F I P P K V H G G 5700
QY 437 -----T L-----438
Db 5701 Y E A Q G E T I E L K I P Y K A Y P Q G E A R W T K G E K I E N N S K F S I T T D D K F A T L R I S N A S R E D Y G 5760
QY 439 -----438
Db 5761 E Y R V V V E N S V G S D S G T V N V T V A D V P P P R P P I E N I L D E A V I L S W K P P A L D G G S L V T N Y T 5820

```
QY 439 -----T-----KH-Q-----E--PV-----445
Db 5821 IEKREMGGSPPCAKSYTYTTIEGLRAGQYEFRIIAENKHGQSKPCEPTAEVLIPGD 5880
QY 446 -----Y-----S-----V-----448
Db 5881 ERKRRGVDVDEQKIVRGKGTVSNDYVFDIWKQYPOPVRIKHDHLDHYDIHEEL 5940
QY 449 -----AF-----S-----P-----452
Db 5941 GTGAGVVHVRTERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAPED 6000
QY 453 D-----G-----R-----Y-----456
Db 6001 DNEWMYIEFMGGELFEKVADEHNKMSDEAVYMRQVCKGLCHMNNYVHLDLKPEN 6060
QY 457 -----L-----A-----458
Db 6061 IMFTTKRSNELKLDLFGTLAHLDPKQSVKVTGTGAFAAPEVABGKPVGYTTDMWSVGL 6120
QY 459 -----SG-S-F-----K-C-----V 466
Db 6121 SYILLGSLPFGGENDDETLRNKVSQDWMNDSDAFSGISBDGKDFIRKLLADPNTRMTI 6180
QY 467 H-----I-----W-----N-----470
Db 6181 HQALEHFWLTPGNAPGRDSQIPSSRYTKIRDSIKTKYDAMPEPLPIGRISNYSRLKHR 6240
QY 471 -----TQ-----G-A--L-----V-H-----478
Db 6241 PQEYSIRDAFWDRSEAPRFIVKPYGTEVGQSANFYCRVIASSPPVVTWHKDDRELKQ 6300
QY 479 S--Y--R-----G-T-----G-I-----FE---488
Db 6301 SVKMKRYNGNDYGLTINRVKDDKGEYTVRAKNSYGTKEEIVFLNVTREHSEPLKFEPL 6360
QY 489 -----V-----C-----W-----491
Db 6361 PMKAPSPRVEFEKERRSAPFFTHLRNRLIQKHOCKLTCSLQGNPNFTIEMKDGHP 6420
QY 492 -----NA-----AG-----D-----K-----V 498
Db 6421 VDEDRVQVFRSGVCSLEIFNARVDDAGTVTATNDLGVDSVCLTVQTKGEPIPRV 6480
QY 499 -----G-----A-----SA-----502
Db 6481 SSFRPRRAYDTLSTGTDVRSYADVRRSLRDVSPDVRSAADDLTKITNELSFSTA 6540
QY 503 --SD-----GS-----V-----507
Db 6541 QLSDSFTEVGSAAEFAASVGSQPEPLIEWLHNGERISEDSRFRASVAGKATLIRISDAK 6600
QY 508 -----C-----VL-----D-----LR--K 514
Db 6601 KSDEGOYLCAASNSAGQEQTRATLTVKGDQPLNGLHAGQAVSELRVTK 6649
```

RESULT 11

S57242
twitchin [similarity] - Caenorhabditis elegans
N;Alternate names: myosin-regulating protein
N;Contains: protein kinase (EC 2.7.1.1)
C;Species: Caenorhabditis elegans
C;Date: 28-Oct-1995 #sequence revision 24-Oct-1997 #text change 09-Jul-2004
C;Accession: S57242; S07571; S06797; S57218; T27934; T28030
submitted to the EMBL Data Library, February 1993
R;Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
A;Description: Additional sequence complexity within twitching of Caenorhabditis elegans
A;Reference number: S57242
A;Accession: S57242
A;Molecule type: DNA
A;Residues: 1-6839 <BEN1>
A;Cross-references: UNIPROT:Q23550; EMBL:L10351

A;Experimental source: var. Bristol
R;Benian, G.
submitted to the EMBL Data Library, November 1989
A;Reference number: S07571
A;Accession: S07571
A;Molecule type: DNA
A;Residues: 792-6839 <BEN2>
A;Cross-references: EMBL:X15423; NID:g6897; PIDN:CAA33463.1; PID:g6898
A;Experimental source: var. Bristol
R;Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.
Nature 342, 45-50, 1989
A;Title: Sequence of an unusually large protein implicated in regulation of myosin activity
A;Reference number: S06797; MUID:9004042; PMID:2812002
A;Accession: S06797
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 806-1175;1178-1998,'Y',2000-3040,'I',3042-3335,'I',3337-5693;5696-6359,'I',6360-6839
A;Cross-references: EMBL:X15423
A;Experimental source: var. Bristol
R;Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
Genetics 134, 1097-1104, 1993
A;Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded protein
A;Reference number: S57218; MUID:93387664; PMID:8397135
A;Accession: S57218
A;Molecule type: DNA
A;Residues: 2-99;108-194,'Q',196-206;374-468;658-753 <BEN4>
A;Experimental source: var. Bristol
R;White, S.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z20442
A;Accession: T27934
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 'MGIPGKKCKQ',19-6839 <WIL>
A;Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CBSP:ZK617.1a
A;Experimental source: clone ZK617
R;Harris, B.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z20458
A;Accession: T28030
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 'MGIPGKKCKQ',19-6839 <WIL2>
A;Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CBSP:ZK617.1a
A;Experimental source: clone ZK829
C;Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
C;Genetics:
A;Gene: unc-22; CBSP:ZK617.1a
A;Map position: 4
A;Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 601/3; 669/3; 676/1; 6808/3
C;Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology; I
C;Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine/threonine kinase
F;806-898,899-990,991-1083,1084-1175,1178-1273,1474-1567,1770-1864,2066-2158,2358-2450,2451-2452,2453-2454,2455-2456,2457-2458,2459-2460,2461-2462,2463-2464,2465-2466,2467-2468,2469-2470,2471-2472,2473-2474,2475-2476,2477-2478,2479-2480,2481-2482,2483-2484,2485-2486,2487-2488,2489-2490,2491-2492,2493-2494,2495-2496,2497-2498,2499-2500,2501-2502,2503-2504,2505-2506,2507-2508,2509-2510,2511-2512,2513-2514,2515-2516,2517-2518,2519-2520,2521-2522,2523-2524,2525-2526,2527-2528,2529-2530,2531-2532,2533-2534,2535-2536,2537-2538,2539-2540,2541-2542,2543-2544,2545-2546,2547-2548,2549-2550,2551-2552,2553-2554,2555-2556,2557-2558,2559-2560,2561-2562,2563-2564,2565-2566,2567-2568,2569-2570,2571-2572,2573-2574,2575-2576,2577-2578,2579-2580,2581-2582,2583-2584,2585-2586,2587-2588,2589-2590,2591-2592,2593-2594,2595-2596,2597-2598,2599-2600,2601-2602,2603-2604,2605-2606,2607-2608,2609-2610,2611-2612,2613-2614,2615-2616,2617-2618,2619-2620,2621-2622,2623-2624,2625-2626,2627-2628,2629-2630,2631-2632,2633-2634,2635-2636,2637-2638,2639-2640,2641-2642,2643-2644,2645-2646,2647-2648,2649-2650,2651-2652,2653-2654,2655-2656,2657-2658,2659-2660,2661-2662,2663-2664,2665-2666,2667-2668,2669-2670,2671-2672,2673-2674,2675-2676,2677-2678,2679-2680,2681-2682,2683-2684,2685-2686,2687-2688,2689-2690,2691-2692,2693-2694,2695-2696,2697-2698,2699-2700,2701-2702,2703-2704,2705-2706,2707-2708,2709-2710,2711-2712,2713-2714,2715-2716,2717-2718,2719-2720,2721-2722,2723-2724,2725-2726,2727-2728,2729-2730,2731-2732,2733-2734,2735-2736,2737-2738,2739-2740,2741-2742,2743-2744,2745-2746,2747-2748,2749-2750,2751-2752,2753-2754,2755-2756,2757-2758,2759-2760,2761-2762,2763-2764,2765-2766,2767-2768,2769-2770,2771-2772,2773-2774,2775-2776,2777-2778,2779-2780,2781-2782,2783-2784,2785-2786,2787-2788,2789-2790,2791-2792,2793-2794,2795-2796,2797-2798,2799-2800,2801-2802,2803-2804,2805-2806,2807-2808,2809-2810,2811-2812,2813-2814,2815-2816,2817-2818,2819-2820,2821-2822,2823-2824,2825-2826,2827-2828,2829-2830,2831-2832,2833-2834,2835-2836,2837-2838,2839-2840,2841-2842,2843-2844,2845-2846,2847-2848,2849-2850,2851-2852,2853-2854,2855-2856,2857-2858,2859-2860,2861-2862,2863-2864,2865-2866,2867-2868,2869-2870,2871-2872,2873-2874,2875-2876,2877-2878,2879-2880,2881-2882,2883-2884,2885-2886,2887-2888,2889-2890,2891-2892,2893-2894,2895-2896,2897-2898,2899-2900,2901-2902,2903-2904,2905-2906,2907-2908,2909-2910,2911-2912,2913-2914,2915-2916,2917-2918,2919-2920,2921-2922,2923-2924,2925-2926,2927-2928,2929-2930,2931-2932,2933-2934,2935-2936,2937-2938,2939-2940,2941-2942,2943-2944,2945-2946,2947-2948,2949-2950,2951-2952,2953-2954,2955-2956,2957-2958,2959-2960,2961-2962,2963-2964,2965-2966,2967-2968,2969-2970,2971-2972,2973-2974,2975-2976,2977-2978,2979-2980,2981-2982,2983-2984,2985-2986,2987-2988,2989-2990,2991-2992,2993-2994,2995-2996,2997-2998,2999-3000,3001-3002,3003-3004,3005-3006,3007-3008,3009-3010,3011-3012,3013-3014,3015-3016,3017-3018,3019-3020,3021-3022,3023-3024,3025-3026,3027-3028,3029-3030,3031-3032,3033-3034,3035-3036,3037-3038,3039-3040,3041-3042,3043-3044,3045-3046,3047-3048,3049-3050,3051-3052,3053-3054,3055-3056,3057-3058,3059-3060,3061-3062,3063-3064,3065-3066,3067-3068,3069-3070,3071-3072,3073-3074,3075-3076,3077-3078,3079-3080,3081-3082,3083-3084,3085-3086,3087-3088,3089-3090,3091-3092,3093-3094,3095-3096,3097-3098,3099-3100,3101-3102,3103-3104,3105-3106,3107-3108,3109-3110,3111-3112,3113-3114,3115-3116,3117-3118,3119-3120,3121-3122,3123-3124,3125-3126,3127-3128,3129-3130,3131-3132,3133-3134,3135-3136,3137-3138,3139-3140,3141-3142,3143-3144,3145-3146,3147-3148,3149-3150,3151-3152,3153-3154,3155-3156,3157-3158,3159-3160,3161-3162,3163-3164,3165-3166,3167-3168,3169-3170,3171-3172,3173-3174,3175-3176,3177-3178,3179-3180,3181-3182,3183-3184,3185-3186,3187-3188,3189-3190,3191-3192,3193-3194,3195-3196,3197-3198,3199-3200,3201-3202,3203-3204,3205-3206,3207-3208,3209-3210,3211-3212,3213-3214,3215-3216,3217-3218,3219-3220,3221-3222,3223-3224,3225-3226,3227-3228,3229-3230,3231-3232,3233-3234,3235-3236,3237-3238,3239-3240,3241-3242,3243-3244,3245-3246,3247-3248,3249-3250,3251-3252,3253-3254,3255-3256,3257-3258,3259-3260,3261-3262,3263-3264,3265-3266,3267-3268,3269-3270,3271-3272,3273-3274,3275-3276,3277-3278,3279-3280,3281-3282,3283-3284,3285-3286,3287-3288,3289-3290,3291-3292,3293-3294,3295-3296,3297-3298,3299-3300,3301-3302,3303-3304,3305-3306,3307-3308,3309-3310,3311-3312,3313-3314,3315-3316,3317-3318,3319-3320,3321-3322,3323-3324,3325-3326,3327-3328,3329-3330,3331-3332,3333-3334,3335-3336,3337-3338,3339-3340,3341-3342,3343-3344,3345-3346,3347-3348,3349-3350,3351-3352,3353-3354,3355-3356,3357-3358,3359-3360,3361-3362,3363-3364,3365-3366,3367-3368,3369-3370,3371-3372,3373-3374,3375-3376,3377-3378,3379-3380,3381-3382,3383-3384,3385-3386,3387-3388,3389-3390,3391-3392,3393-3394,3395-3396,3397-3398,3399-3400,3401-3402,3403-3404,3405-3406,3407-3408,3409-3410,3411-3412,3413-3414,3415-3416,3417-3418,3419-3420,3421-3422,3423-3424,3425-3426,3427-3428,3429-3430,3431-3432,3433-3434,3435-3436,3437-3438,3439-3440,3441-3442,3443-3444,3445-3446,3447-3448,3449-3450,3451-3452,3453-3454,3455-3456,3457-3458,3459-3460,3461-3462,3463-3464,3465-3466,3467-3468,3469-3470,3471-3472,3473-3474,3475-3476,3477-3478,3479-3480,3481-3482,3483-3484,3485-3486,3487-3488,3489-3490,3491-3492,3493-3494,3495-3496,3497-3498,3499-3500,3501-3502,3503-3504,3505-3506,3507-3508,3509-3510,3511-3512,3513-3514,3515-3516,3517-3518,3519-3520,3521-3522,3523-3524,3525-3526,3527-3528,3529-3530,3531-3532,3533-3534,3535-3536,3537-3538,3539-3540,3541-3542,3543-3544,3545-3546,3547-3548,3549-3550,3551-3552,3553-3554,3555-3556,3557-3558,3559-3560,3561-3562,3563-3564,3565-3566,3567-3568,3569-3570,3571-3572,3573-3574,3575-3576,3577-3578,3579-3580,3581-3582,3583-3584,3585-3586,3587-3588,3589-3590,3591-3592,3593-3594,3595-3596,3597-3598,3599-3600,3601-3602,3603-3604,3605-3606,3607-3608,3609-3610,3611-3612,3613-3614,3615-3616,3617-3618,3619-3620,3621-3622,3623-3624,3625-3626,3627-3628,3629-3630,3631-3632,3633-3634,3635-3636,3637-3638,3639-3640,3641-3642,3643-3644,3645-3646,3647-3648,3649-3650,3651-3652,3653-3654,3655-3656,3657-3658,3659-3660,3661-3662,3663-3664,3665-3666,3667-3668,3669-3670,3671-3672,3673-3674,3675-3676,3677-3678,3679-3680,3681-3682,3683-3684,3685-3686,3687-3688,3689-3690,3691-3692,3693-3694,3695-3696,3697-3698,3699-3700,3701-3702,3703-3704,3705-3706,3707-3708,3709-3710,3711-3712,3713-3714,3715-3716,3717-3718,3719-3720,3721-3722,3723-3724,3725-3726,3727-3728,3729-3730,3731-3732,3733-3734,3735-3736,3737-3738,3739-3740,3741-3742,3743-3744,3745-3746,3747-3748,3749-3750,3751-3752,3753-3754,3755-3756,3757-3758,3759-3760,3761-3762,3763-3764,3765-3766,3767-3768,3769-3770,3771-3772,3773-3774,3775-3776,3777-3778,3779-3780,3781-3782,3783-3784,3785-3786,3787-3788,3789-3790,3791-3792,3793-3794,3795-3796,3797-3798,3799-3800,3801-3802,3803-3804,3805-3806,3807-3808,3809-3810,3811-3812,3813-3814,3815-3816,3817-3818,3819-3820,3821-3822,3823-3824,3825-3826,3827-3828,3829-3830,3831-3832,3833-3834,3835-3836,3837-3838,3839-3840,3841-3842,3843-3844,3845-3846,3847-3848,3849-3850,3851-3852,3853-3854,3855-3856,3857-3858,3859-3860,3861-3862,3863-3864,3865-3866,3867-3868,3869-3870,3871-3872,3873-3874,3875-3876,3877-3878,3879-3880,3881-3882,3883-3884,3885-3886,3887-3888,3889-3890,3891-3892,3893-3894,3895-3896,3897-3898,3899-3900,3901-3902,3903-3904,3905-3906,3907-3908,3909-3910,3911-3912,3913-3914,3915-3916,3917-3918,3919-3920,3921-3922,3923-3924,3925-3926,3927-3928,3929-3930,3931-3932,3933-3934,3935-3936,3937-3938,3939-3940,3941-3942,3943-3944,3945-3946,3947-3948,3949-3950,3951-3952,3953-3954,3955-3956,3957-3958,3959-3960,3961-3962,3963-3964,3965-3966,3967-3968,3969-3970,3971-3972,3973-3974,3975-3976,3977-3978,3979-3980,3981-3982,3983-3984,3985-3986,3987-3988,3989-3990,3991-3992,3993-3994,3995-3996,3997-3998,3999-4000,4001-4002,4003-4004,4005-4006,4007-4008,4009-4010,4011-4012,4013-4014,4015-4016,4017-4018,4019-4020,4021-4022,4023-4024,4025-4026,4027-4028,4029-4030,4031-4032,4033-4034,4035-4036,4037-4038,4039-4040,4041-4042,4043-4044,4045-4046,4047-4048,4049-4050,4051-4052,4053-4054,4055-4056,4057-4058,4059-4060,4061-4062,4063-4064,4065-4066,4067-4068,4069-4070,4071-4072,4073-4074,4075-4076,4077-4078,4079-4080,4081-4082,4083-4084,4085-4086,4087-4088,4089-4090,4091-4092,4093-4094,4095-4096,4097-4098,4099-4100,4101-4102,4103-4104,4105-4106,4107-4108,4109-4110,4111-4112,4113-4114,4115-4116,4117-4118,4119-4120,4121-4122,4123-4124,4125-4126,4127-4128,4129-4130,4131-4132,4133-4134,4135-4136,4137-4138,4139-4140,4141-4142,4143-4144,4145-4146,4147-4148,4149-4150,4151-4152,4153-4154,4155-4156,4157-4158,4159-4160,4161-4162,4163-4164,4165-4166,4167-4168,4169-4170,4171-4172,4173-4174,4175-4176,4177-4178,4179-4180,4181-4182,4183-4184,4185-4186,4187-4188,4189-4190,4191-4192,4193-4194,4195-4196,4197-4198,4199-4200,4201-4202,4203-4204,4205-4206,4207-4208,4209-4210,4211-4212,4213-4214,4215-4216,4217-4218,4219-4220,4221-4222,4223-4224,4225-4226,4227-4228,4229-4230,4231-4232,4233-4234,4235-4236,4237-4238,4239-4240,4241-4242,4243-4244,4245-4246,4247-4248,4249-4250,4251-4252,4253-4254,4255-4256,4257-4258,4259-4260,4261-4262,4263-4264,4265-4266,4267-4268,4269-4270,4271-4272,4273-4274,4275-4276,4277-4278,4279-4280,4281-4282,4283-4284,4285-4286,4287-4288,4289-4290,4291-4292,4293-4294,4295-4296,4297-4298,4299-4300,4301-4302,4303-4304,4305-4306,4307-4308,4309-4310,4311-4312,4313-4314,4315-4316,4317-4318,4319-4320,4321-4322,4323-4324,4325-4326,4327-4328,4329-4330,4331-4332,4333-4334,4335-4336,4337-4338,4339-4340,4341-4342,4343-4344,4345-4346,4347-4348,4349-4350,4351-4352,4353-4354,4355-4356,4357-4358,4359-4360,4361-4362,4363-4364,4365-4366,4367-4368,4369-4370,4371-4372,4373-4374,4375-4376,4377-4378,4379-4380,4381-4382,4383-4384,4385-4386,4387-4388,438

QY 12 --VYR-----Y-----L-----16
Db 912 QEIYRTQAVLTCKVNTPRAPLVYRGSKAIQEGDPRFIEIKDAVGRCTLTIKEVEDQ 971
QY 17 -----Q-----ES-----19
Db 972 AEWTAIRITQDVFSKVQVYVEPRHTFVPMKSKQVNESDLATLETVDNDKDAEVMWHDG 1031
QY 20 --G--F--S-----H-----S-----A--F--26
Db 1032 KRIDIDGVAFKVBESSNRKRRLIINGARIEDHGEYKCTTKDDRTMAQLIIVDAKKNFIVALK 1091
QY 27 -----T-----F--G--I-----K-----S-----H-----33
Db 1092 DTEVIEKDDVTLMQTKDTKPGIWFNRGKQISSMPGKFETQSRNGTHTLKIGKIEMNE 1151
QY 34 ----ISO-----S-----N-----38
Db 1152 ADVYEDQAGLRGSCNVTLEAEKRPILNWKPKKIEAKAGPCVVKVPFQIKGTRRGDPK 1211
QY 39 --I--NG-----N-----AL-----43
Db 1212 AQLKNGKPIDEMRKLVEVIIKDDVAEIVFKNPQLADTGKWALELGNAGTALAPFELF 1271
QY 44 V-----PP-----AA-----LI--S-----II--Q-----KG 56
Db 1272 VKDKPKPKPLTKNVTABGLDLVNGTPDPDEGAPVKAYIIEMQGRSGNMAKVGETKG 1331
QY 57 -----L-----QY-----VEA-----E--V-----SI-----N-----67
Db 1332 TDFVKVLDKEHGEYKFRVKALNEGLSLPTGESVLAKNPYGVPGPKQMDAIDVDKDH 1391
QY 68 -----EDG-----T-----L-----72
Db 1392 TLAWPEPEDGGAPITGYIIERREKSEKWHQVQGTQKPCCELTDKKVVEDKEYLYRVKA 1451
QY 73 -----F-----D-----74
Db 1452 VNKAGPGDPCDHGKPIKMAKAKASPEFTGGIKDLRLKVGETIKYDVPISGEPLPECLW 1511
QY 75 --GRP-----IE-----S-----L-----S-----82
Db 1512 VNGKPLKAVGRVKMSSERGHKMIENAVRADSGKFTITLKNSSGSCDSTATVTVVGRPT 1571
QY 83 -----L--I-----D--AVM-----PD-----VV--Q-----T--94
Db 1572 PPKGPLDIADVADGATLSWNPPDDGDPPLTGYIIVEAQMDNKGKIIIEVGKVDNPTTL 1631
QY 95 -----R-----Q-----Q-----97
Db 1632 KVNGLRNGNYKFRVKAVNNEGESEPLSADQYTIKDPWDEPGKPGRPEITDFADRID 1691
QY 98 -----A-----Y-----RD-----101
Db 1692 AWEPPHKGGAPEEYIEVVRDPTKEMKEVRVPDNTNASISGLKEGKEYQFRVAVNKA 1751
QY 102 -----K--LA-----Q-----105
Db 1752 GPCQSPSEKQLAKPKFIPAWLKHDNLKSIITVKAGATVRWEVKIGGEPIPEVKWPKGNQ 1811
QY 106 Q-----Q-----A-----A-----A-----110
Db 1812 QLENGIQLTIDTRKNEHTILCIPSAKRSVDGEYRLTVKNSHGADESKANLTVLDRPSKN 1871
QY 111 -----A-----A-----A-----A-----113
Db 1872 GPLEVDVPEDNLNLSWKPPDDGGEPIEYIEVEKLDATGRVWPCAKVKDKTAKHIDGLK 1931
QY 114 -----A-----A-----A-----A-----118
Db 1932 KGQTYQFRVKAVNKEGASDALSTDKTKAKNPYDEFGKTGTDPDWDADADRSLEWEPK 1991
QY 119 S-----Q-----Q--G--S-----A-----K--N-----GE 128

Db 1992 SDGGAPITQYVIEKKGKGRDWEQCKVSGDQTNABILGLKEGEYQFRVKAVNKA 2051
QY 129 -----N-----T--AN-----CE-----E--N-----136
Db 2052 ASDPSKRVKAPRNLPKWDREAMKTITIKVGNDFEFDVPRGEPKKEWIFNEKPVDD 2111
QY 137 -----GA--H-----TI--A--N-----TDM-----ME--150
Db 2112 QXIRIBESDYKTRFVLRGATRKHAGLYLTATNASGDKHSVEVILGKPSPLGLEVS 2171
QY 144 N-----H-----TDM-----150
Db 2172 NVYEDRADLEWKVPEDGGAPIDHYIEKMDLATGRWVPCGRSETTKTTVPNLQPGHEYK 2231
QY 151 -----V-----D-----G-----D-----V 155
Db 2232 FRVRANKGESDPLTTNTAILAKNPYEPGVKVDKPELVDWKDHDVLDLAWNAPDDGAPI 2291
QY 156 E--I-----P-----P-----NKA-----V 163
Db 2292 EAFVIEKKDKNGRWEALVPGDQKTATVPNLKEGEYQFRISARNKAGTGDPSPDRV 2351
QY 164 V-----L-----R-----G-----H-----ES-----170
Db 2352 VAKPRNLAPRIHREDLSDTTVKVGATLKFIHVHIDGEPAPDVTWFSNKGKIGESKAQIENE 2411
QY 171 -----171-----EV-----172
Db 2412 PYISRPALPKARKOSGKYTITATNINGTSDVTINIKVSKPKPKGPIEVTDFEDRAT 2471
QY 173 -----F-----I--C-----175
Db 2472 LDKPPEDDGGPIEYIEKMKTKDGIWVPCGRSGDTHFTVDSLNGDHYKFRVKA 2531
QY 176 -----176-----A-----176
Db 2532 EGSPDPLETETDILAKNPDRPDRPGRPEPTDWDSDHVDLKWDPPLSDGGAPIEYQIEK 2591
QY 177 -----W-----177
Db 2592 RTKYGWEPATITVPGGQTATVPDLTPNBEYFRVAVNKGSPSPDASKAVIAPRN 2651
QY 178 -----N-----PVS-----D-----L-----183
Db 2652 KPHIDRALKNLTIKAGQISFDVPVSGEPAPTIVHWPDRNREIRNGRVLKDNPEYQSK 2711
QY 184 L-----A-----S--G-----S--GD-----190
Db 2712 LVVKQMERGDSGTFTIKAVNANGEDATVKINVIDKPTSPNGPLDSDVHGDHVTLNWRA 2771
QY 191 -----S-----TA-----R-----194
Db 2772 PDDGGIPIENYVIEKYDTASGRWVPAKAVAGDKTTAVVDGLIPGHEYKFRVAAVNA 2831
QY 195 -----195-----I-----195
Db 2832 SDPLETFGTTLAKDPDKGKTNAPEITDWDKDHVDLEWKPANDGAPIEEYVEMKDE 2891
QY 196 -----W-----NL--S--E-----200
Db 2892 FSPFMDVAHVPAQTNAVTGNLKEGSKYEFIRAKNKAAGLGDPSDSASAKARNVPPV 2951
QY 201 ---NS-----T--S-----GST-----207
Db 2952 IDRNSIQIBIKVKAQDFSLNI PVSGEPTTITWTPEGTPEVSDRMKLANNEDGKTFHVK 3011
QY 208 Q--L-----VL-----R-----212
Db 3012 RALRSTGTGYIIKAENENGTDAEVKVTVDHPSSPRGRLDVTNIVKDGCDLAWKEPDD 3071
QY 213 -----H-----C-----214

Db 3072 GBAISHYVIEKODAAATGRWTACGESKDTNFHVDLTAQGHEYKFRVAVNRHGDSDPLEA 3131
Qy 215 ---I---R---EG--- 218
Db 3132 REALIAKDPDRADKPGTPELVMDKDHADLKWTTPADDDGAGPIEGYLVEMRTPSGDWVP 3191
Qy 219 --- 218
Db 3192 AVTVGAGELTATVDGLKPGQTYQFRVKALNKAGESTSPDSRTWVAKPRHLAKINRDMF 3251
Qy 219 ---GO---DV---P--- 223
Db 3252 VAORVAGOTLNFVNVEGEPAPKIEWFLNGSPILSSGCGNTHIDNNTDNNTKLTAKSTARA 3311
Qy 224 ---S---N---KDV---S---LD--- 232
Db 3312 DSGYKIVATNESKDBHEVDVNDILDPGAPGELRHKIDTKGSVILKWDEPLDDGGSPI 3371
Qy 233 ---W---N---SE---CT---L--- 239
Db 3372 TNYVVEKQEDGGRWVPCGETSDTSLKYNKLSGHEYKFRVAVNRQTSAPLSDHAIVA 3431
Qy 240 ---L---A---T---G---SY---DG 247
Db 3432 KNPFEDEPAPTDTVPDWDKDHVLEWKPANDGGAPIDAYIIVEKKDKFGDWVECARVDG 3491
Qy 248 ---F---A---RI--- 251
Db 3492 KTKATADNLTPGETYQFRVAVNKAGPGKPSDPTGNVAVKPRMAPKLNLAGLLDIRK 3551
Qy 252 ---W---TKD---G---NL---A---S---T---L--- 262
Db 3552 AGTPIKLDIAFEGEPAPVAKKANDATIDTGARADVTNTTSSAIHIFSAVRGDTGVYKI 3611
Qy 263 ---G---Q---HK---G---P---I--- 269
Db 3612 IVENEHGKTAQCNVTVDVPGTPEGPKIDEIHKEGCTLNWKPPTDNGGTDVLHVIVEK 3671
Qy 270 ---F---A---L--- 272
Db 3672 MDSRGTWQEVGTPDCTAKNKLVPGEYAFRVAFAVKNVLOGESKPLEAEPIIAKQFDV 3731
Qy 273 ---KWN---KK---G--- 278
Db 3732 PDPVKPEVTDWDKORIDIKNWPNANNGAPVTGYIIVEKKEKSAIWEAGKPTGTTFA 3791
Qy 279 -N---F---IL---S---AG--- 285
Db 3792 DNLKPGVEYFRVIAVNAAGSPDPTDPQITKARYLKPILKTASRKIKIKAGFTHNLE 3851
Qy 286 ---VD---K---TT--- 290
Db 3852 VDFIGAPDPTATWVGSGAALAPELLVDKASSTTSIFPFSAKRADSGNYKLVKNELGE 3911
Qy 291 ---I---I---W---D--- 294
Db 3912 DEAFIVIQDRSPAGPLEVSDVTKDSCVLNWKPKPDGGAEISNYVVEKEDTKNTW 3971
Qy 295 ---A---H---T---G---E---AKQOF--- 304
Db 3972 VPVSFAVTGTSITVPKLTEGHEYFRVMAENTGCRSDSLNTDFVLAKDPFGTPGKGRP 4031
Qy 305 ---P---F--- 306
Db 4032 EIVDTDNHDHIDIKWPPRDNGSGPVHDYDIERKDAKTGRWIKNTSPVQGTAFSDTRVQK 4091
Qy 307 -H---SA---P---A---LD---V--- 314
Db 4092 GHTYEYRVAVNAKAGCPQSDSSAAATAKPMHEAPKFDLDDGKEFRVKAPELVITIPF 4151
Qy 315 ---D---W---Q---S--- 318
Db 4152 TASPQPDISWTKEGKPLAGVETTDSTQTKLVIPSTRSDSGPVKIKAVNPYGEAEANIKI 4211

Qy 319 ---N---N--- 320
Db 4212 TVIDKCAPENITYPAVSRHTCTLNWDAPKDDGAGIAGYKIEYQEVGSOIWDKVPGLIS 4271
Qy 321 ---T---F---A---S---C---ST--- 327
Db 4272 GTAYTVRGLEHGQYFRIRAEANAVGLSDYCOQVPVVIKPPDPGAPSPTEITGYDTNQ 4331
Qy 328 ---D---M---CI--- 331
Db 4332 VSLAMPNRDDGGSPILGYVVERFEKRGGDWAPVXMPVMVKGTCEICVPLHENETYQFRV 4391
Qy 332 ---H---V---C--- 334
Db 4392 RAVNAACHGEPNSGSEPVTCRPVVEKPGADAPRVGKITKNSAELTWRNRLRGGAPIDG 4451
Qy 335 ---KLK--- 337
Db 4452 YIVEKKLGDNDWTRCNDKPVRTAFEVKNLGEKEEYFRVIAVNSAGEGEPKPSDLVL 4511
Qy 338 ---Q---D---R---PI--- 342
Db 4512 IEEQGRPIFDINNLDKITVRAGETIQIRIPYAGGNPKPIIDLFGNNGSPIFENERTVVDV 4571
Qy 343 ---K---T---F---Q--- 346
Db 4572 NPGEIVITTTGSKSDAGPYKISATNKYKDKTCKLVFVLDAPKPTGPIRATDIQADAM 4631
Qy 347 ---GH---T---N---E---V---N--- 353
Db 4632 TLSWRPPKONGDAITNYVVEKRTPGGDWVTGHPVGTTLRVNLDANTPEYFRVAENQ 4691
Qy 354 ---AI---K---W---D---P--- 359
Db 4692 YGVGEPLETDDAIVAKNPPFTPGAQOPEAVETSEBAILTQWTRPTSDGAPIQGVYIEK 4751
Qy 360 ---T---GN---L--- 362
Db 4752 REVGSTEWTKAAGFNILDTKHRVTGLTPKTYEFRVAAVNAAGQGEYSVNSVPIADNAP 4811
Qy 363 ---L--- 363
Db 4812 TRPKINMGMLTRDILAYAGERAKILVFAASPAPKVFSGKNGKISPTDPRVKVEYSDFL 4871
Qy 364 ---L---AS---C--- 367
Db 4872 ATLTIKSELTDGGLYFVLENSQGSDSASIRLKVVDKSPASPOHIRVEDIAPCCTLYWM 4931
Qy 368 ---SD---DM---T--- 372
Db 4932 PPSDDGSPITNYIVEKLDLRHSDGKWEKVSFVRNLNTVGGLIKDNRYRFRVRAETQY 4991
Qy 373 ---L---K---I---W---SMK---Q--- 380
Db 4992 GVSEPCELADVVKVQYFVFNQPEAPTVRDKDSTWAELEWDPRDGS-KIIGYQVQYR 5050
Qy 381 D---N---CVHD---L---Q--- 389
Db 5051 DTSSGRWINAKMDLSEQC-HARVTGLRQNGEPEFRIIAKNAAGFSKPSPPERCOLKSFR 5109
Qy 390 ---H---N---K---EI---Y---TIKW---S--- 400
Db 5110 GPPGPIHVGAKSIGNHCTITWMAPLEDGGSKITGYNVREIYEGSTL-WTVASDYNVRE 5168
Qy 401 P---T---GP---G---T---N--- 407
Db 5169 PEFTVDKLRFNDEYFRVAVNAAGKIPSLPSGPIKIQESGGRRPQIVVVKPEDTAQPN 5228
Qy 408 ---N---P--- 409
Db 5229 RRAVFTCEAVGREPTARWLNRNGRELPESSRYRFEASDGYKFTIKEVMDIDAGEYTV 5288

QY 410 -N-----ANL----- 413
Db 5289 SNFYSGDTATANLVQAPPVIEKDVNTILPSGLVRLKIYFSGTAPFRHSLVNLNEED 5348
QY 414 M-----L-----A-S-A-----S-----F----- 420
Db 5349 MDHPTIRIVEFDHILITIPALSVREAGRYEYTVSNDSGEATTFGLNVTGLPEAPQGPL 5408
QY 421 -----D-S-----TV----- 424
Db 5409 HISNIGSTATLSWRPVTDDGSKITSYVVEKEDLSKDEWVTYSNVKDMYIIVTGLPEN 5468
QY 425 -----W----- 424
Db 5469 HEYEFVRSQENIGAPLVSEHPHIIARLPDPPTPLNLEIVQVGGDYVTLWSQRPLSD 5528
QY 425 -----RL----- 427
Db 5529 GGGRLRGYIVKEQEBEHDFRCNQNPSPNNVNPNLIDGRKYRYRFAVNDAGLSDLA 5588
QY 428 --D-----V-----D-----R----- 431
Db 5589 ELDOTLQASGSGEGPKIVSPLSDLANEEVRCVTFECEISGSPRPRPYRWFKGCKELVDTS 5648
QY 432 -----G-----I-----C-----I-----H----- 436
Db 5649 KYTLINKGDKQVLIINDLTSDDADEVTCRATNSSGTRSTRANLRITKTPRVFIPPKYHGG 5708
QY 437 -----Y-----S-----TL----- 438
Db 5709 YEAQGETIELKIPYKAYPQGEARWTKGGEKIENNSKFSITDDKFPATLRISNASREDYG 5768
QY 439 -----T-----KH-Q-----E-----PV----- 445
Db 5829 IEKREAMGSGWSPCAKSRVYTTIEGLRAGQVYFPIIAENKHGQSKPCEPTAPVLIPGD 5888
QY 446 -----Y-----S-----V----- 448
Db 5889 ERKRRRGYDVEQGIKVRGKTGVSSNYDNYVDFIWKQYYPQVPEIKHVDLHYDIHEEL 5948
QY 449 ---AF-----S-----P----- 452
Db 5949 GTCAFGVHRVTERATGNPAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAFED 6008
QY 453 D-----G-----R-----Y----- 456
Db 6009 DNEWMVIEBFMSGGLFEKVADEHNKMSDEAVEYMRQVCKGLCHMHENNYVHLDLKPEN 6068
QY 457 -----L-----A----- 458
Db 6069 IMFTTKRSNELKLIDFGLTAHLDPKOSVKVTTGTAEFAAPEVAEGKPVGYTDMWSVGL 6128
QY 459 ---SG-S-F-----D-----K-C----- 466
Db 6129 SVILLSGLSPFGGNDDETLRVKSCDWNDDSAFSGISGDKDFTRKLLADPNTRMTI 6188
QY 467 H-----I-----W-----N----- 470
Db 6189 HQALEHPMLTPGNAPGRDQIPSRVTKIRDSIKTKYDAWPEPLPLGRISNVSSLRKH 6248
QY 471 -----TQ-----T-----G-A-L-----V-H----- 478
Db 6249 POEYSIRDADFDRSEAPRFIVKPYGTEVGEOSANFYCRVIASSPPVVTWHKDDRELKQ 6308
QY 479 S-Y-R-----G-T-----G-----I-----PE----- 488
Db 6309 SVKMKRYNGNDYGLTINRVKGDGKEYTVRAKNYSYGTKEEIVFLNVTNRHSBPLKPEPLE 6368
QY 489 -----V-----C-----W----- 491

Db 6369 PMKAPSPRRVEBEFKERRSGAPFFTHLRNRLIQKNHCKLCTSLQGNPNPTIEWMKDGH 6428
QY 492 -----NA-----AG-----D-----K-----V 498
Db 6429 VDEDRVQVFRSGVCSLEIFNARVDDAGTYVTATNDLGVVDVSECVLTVQTKGGEIPRV 6488
QY 499 -----G-----A-----SA----- 502
Db 6489 SSFPRRAYDTLSTGTDVDRSHSYADMRRSLRDVSPDVSAAADDLKTITNELPSFTA 6548
QY 503 --SD-----GS-----V----- 507
Db 6549 QLSDETSEVGSAAEFAAAGVGOPELIEWLHNGERISEDSRFRASVYAGKATLRISDAK 6608
QY 508 -----C-----VL-----D-----LR--K 514
Db 6609 KSDEGYLCRASAGQEOTRATLTVKGDQPLNHHAGQAVESELRTVK 6657

RESULT 12

T27935

hypothetical protein ZK617.1b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T27935; T28031

R:White, S.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z20442

A:Accession: T27935

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-7160 <W1>

A:Cross-references: UNIPROT:Q23551; EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617.1b

A:Experimental source: clone ZK617

R:Harris, B.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z20458

A:Accession: T28031

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-7160 <W1>

A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b

A:Experimental source: clone ZK829

C:Genetics:

A:Gene: CESP:ZK617.1b

A:Map position: 4

A:Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 599/3

A:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology; I

Query Match 70.3%; Score 2572.2; DB 2; Length 7160;

Best Local Similarity 8.0%; Pred. No. 5.6e-52;

Matches 472; Conservative 29; Mismatches 10; Indels 5358; Gaps 394;

QY 1 M-----S-----I-----S-S-----D----- 6
Db 1113 MTEAERRQLFPGKKVEKWDIPLPEKTVQOQVDKICEWKCTYSRPNKIRWYKDKRKEIFS 1172
QY 7 -----EV-----N-----PL----- 11
Db 1173 GGLKYKIVIEKNVCTLIINNPEVDOTGKYTCBANGVPTHAQLTVLEPPMKYSPLNPUNT 1232
QY 12 --VYR-----Y-----L----- 16
Db 1233 QBIYRTKQAVLTKVNTPRAPLVWYRGSKAIQEGDPRFIEKDAVGRCTLTITKEVEDDQ 1292
QY 17 -----Q-----ES----- 19
Db 1293 AEWTAIRITQDVFQVQVVEPRHPTFVPMKSCQKNESDLATLETVDNDKDAEIVVWHG 1352
QY 20 -----G-F-----S-----H-----A-----F----- 26

Db 1353 KRIDIDGVKFKVSSNRKRLIINGARIEDHGEYKCTTKDRTMAQLIVDAKNKFI VALK 1412
QY 27 -----T-----F-G-I-----K-----S-----H----- 33
Db 1413 DTEVIEKDDVTLMCQTKDTKPGIWFNRNGKOISSMPGCKFTQSRNGHTLTKIGKIEWNE 1472
QY 34 -----ISQ-----S-----N----- 38
Db 1473 ADVVEIDQAGLRGSCNVTVLEAEKRPILNWKPKKIEAKAGEPCVVKVFFQIKGTRRGDPK 1532
QY 39 -----I-----NG-----AL----- 43
Db 1533 AQILKNGKPIDEEMRKLVEIHKDDVAEIVFKNPQLADTKGWALELNGSAGTALAPPFLF 1592
QY 44 V-----PP-----AA-----LI-----S-----II-----Q-----KG 56
Db 1593 VKDKPKPKPLETKNVTAEGLDLVWGTPDDPEGAPVKAVIIEOGRSGNWAQVGETKG 1652
QY 57 -----L-----QY-----VEA-----E-V-----SI-----N----- 67
Db 1653 TDFKVKDLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPYGVPGKPKNMDAIDVDKXHC 1712
QY 68 -----EDG-----T-----L----- 72
Db 1713 TLAWEPPEEDGAPITGVIIERREKSEKWHQVGTQKPCCELTDKKVEDKEYLYRVKA 1772
QY 73 -----F-----D----- 74
Db 1773 VNKAQGDPCDHGPKPKAKKAKASPEFTGGIKDLRLKVGETIKYDVPISGEPLPECLMW 1832
QY 75 -----GRP-----IE-----S-----L-----S----- 82
Db 1833 VNGKPLKAVGRVMSSEKHKIMKIENAVRADSKFTITLKNSSGSCDSTATVTVGRPT 1892
QY 83 -----L-I-----D-AWM-----PD-----VW-Q-----T- 94
Db 1893 PPKGFLDIADVADGATLSWNPDDDGCDPLTGVIEAODMNDKNGKIEVGVKVDPNNTTL 1952
QY 95 -----R-----Q-Q----- 97
Db 1953 KYNGLRNKGNVFRVKAVNNEGESEPLSADQYQIKDPWDEPGKPRPEITDFADRIDI 2012
QY 98 -----A-----Y-----RD----- 101
Db 2013 AWEPPHKDGGAPIEYIEYVVRDPTDKWKEVKRVDPDTNASISGLKEGKEYQFRVAVNKA 2072
QY 102 -----K-LA----- 105
Db 2073 GPGQSPSEPKLAKPKFIPAWLKHDLNLSITVKAGATVRWEVKIGGEPPEVVKWFKGNQ 2132
QY 106 Q-----Q-----A-----A-----A----- 110
Db 2133 QLENGIQLTIDTRKNEHTILCIPSMRSDVGEYRLTVKNSHGADBEKANLTVLDRPSKN 2192
QY 111 -----A-----A-----A-----A-----A----- 113
Db 2193 GPLEVSDVFDNLNLSWKPPDDGGEPYIEYVEKLDJTAGRWVPCAKVKDTKAHIDGLK 2252
QY 114 -----A-----A-----A-----A-----A----- 118
Db 2253 KGQTYQFRVKAVNKEGASDALSTDKTKAKNPYDEFGKTGTGTPDWDMDADRVSLWEPPK 2312
QY 119 S-----Q-----Q-G-S-----A-----K-N-----GE 128
Db 2313 SDGAPITQYVIEKKKGKHDWQECGVKSGDQTNAEILGLKEGEYQFRVAVNKAQGPGE 2372
QY 129 -----N-----T-----AN-----GE-----E-N----- 136
Db 2373 ASDPSRKVAKPNLKPWIDREAMKTTIKVGNDFVDFVVRGEPPEPKKEWIFNEKPVDD 2432
QY 137 -----GA-----H-----TI-A-N----- 143
Db 2433 QKTRIESEDYKTRFVLRGATRKHAGLYTLTATNASGSDKHSVEVIVLGRPSSPLGPLEVS 2492

QY 144 N-----H-----TDM-----ME----- 150
Db 2493 NVYEDRADLEWKVPEDDGGAPIDHYIEKMDLATGRWVPCGRSETTKTTVNLNLPQGHYK 2552
QY 151 -----V-----D-----G-----D-----V 155
Db 2553 FRVAVNKEGESDPLTNTAILAKNPYEVPGKVDKPELVWDKXHDVLDLAWNAPDDGAPI 2612
QY 156 E-----I-----P-----P-----NKA-----V 163
Db 2613 EAFVIEKKDKNGRWEALVVPDQKTATVNLKEGEYQFRSARNKAGTGDPSPDRV 2672
QY 164 V-----J-----R-----G-----H-----ES----- 170
Db 2673 VAKPRNLAPRIHREDLSDTTIVKVGATLKFIVHIDGEPAPDVTWSPNGKIGESKAQIENE 2732
QY 171 ----- 172
Db 2733 PYISRFPALPKALRKQSGKYTITATNINGDTSVTINIKVSKPTKPKGPIEVTDFEDRAT 2792
QY 173 -----F-----I-----C----- 175
Db 2793 LDWKPPEDDGGEPYIEFYIEKMTKDIWVPCGRSGDTHFTVDSLNGDHYKFRVKA VNS 2852
QY 176 ----- 176
Db 2853 EGSPDPLETETDILAKNPFRDRPDRPGRPEPTDWDSDHVDLKWDPPLSDGGAPIEYQIEK 2912
QY 177 -----W----- 177
Db 2913 RTKYGRWEPAITVPGGQTTATVPDLTPNEEYFRVAVNKGSPSPSDASKAVIAKPRNL 2972
QY 178 -----N-----PVS-----D-----L----- 183
Db 2973 KPHIDRALKNLTIKAGOSISFDPVPSGEPAPTWTWHWPNDRNREIRNGRVLKDNPEYQSK 3032
QY 184 L-----A-----S-G-----S-GD----- 190
Db 3033 LVVKQMERGDSGFTTIKAVNAGEDATVKINVDKPTSPNGPLDVSDVHGDHVTUNWRA 3092
QY 191 -----S-----TA-----R----- 194
Db 3093 PDDGGIPIENVVIEKYDTASGRWVPAKAVAGDKTTAVVDGLIPGHEYKFRVAAVNAEGE 3152
QY 195 ----- 195
Db 3153 SDPLETFGTTLAKDPDFKGTNAPEITDWDKXHDVLEWKPPANDGGAPIEYVVEWKDE 3212
QY 196 -----W-----NL-----S-E----- 200
Db 3213 FSPFNDVAVHVPAGQTNATVGNLKEGSKYEFIRAKNKAAGLGDPSDSASAVAKARNVPPV 3272
QY 201 -----NS-----T-----S-----GST----- 207
Db 3273 IDRNSIQEIKVKAGQDFSLNIPVSGEPTPTITWTPEGTPEVSDDRMKNLNNEDGKTFHKV 3332
QY 208 Q-L-----VL-----R----- 212
Db 3333 RALRSDTGTYYIIKAENENGTDTAEVKVTLDHPSSPRGPLDVTNIVKDGCDLAWKEPEDD 3392
QY 213 -----H-----C----- 214
Db 3393 GGAESHVYIEKQDAATGRWTACGESKOTNFHVDDLTOGHEYKFRVAVNRHGDSDPLEA 3452
QY 215 -----I-----R-----EG----- 218
Db 3453 REAILAKDPFDRADKPGTPEIVDWDKXHDKWTTPADDGGAPIEGYLVEWMTSPSGDWVP 3512
QY 219 ----- 218
Db 3513 AVTVGAGELTATVDGLKPGQTYQFRVKALNKAGESTPSPSRMTWAKPRHLAPKINRDMF 3572

QY 219 -----GQ-----DV-----P----- 223
Db 3573 VAQVRKAGQTLNFDVNVGEPAKPIBWFLNGSLSSGGNTHIDNNTDNNKLTTKSTARA 3632
QY 224 -----S-----N-----KQVT--S-----LD----- 232
Db 3633 DSGKYKIVATNESGKDEHEVDNILDIPGAPEGLRHKDITKESVVLKWDPELDGSGSPI 3692
QY 233 -----W-----N-----SE-----GT--L----- 239
Db 3693 TNYVVEKQEDGGRVPCGETSDTSLKVNKLSEGEHYKFRKAVNRQGTSAPLTSDHAIYA 3752
QY 240 -----L-----A-----T-----G-----SY-----DG 247
Db 3753 KNPFDEPDAPTDVTPDWDKDHVLEWKPPANDGGAPIDAYIVEKDKFGDMWECARVDG 3812
QY 248 -----F-----F-----A-----RI-- 251
Db 3813 KTTKATADNLTPGETTYQFRKAVKAGPKGKPSDPTGNVAKPRMAPKLNLAGLLDLRIK 3872
QY 252 -----W-----TKD--G-----NL-----A-----S-----T--L-- 262
Db 3873 AGTPIKLDIAFEGEPAPVAKWKANDATIDTGARADVTNTPTSSAIHFSAVRGDTGVYKI 3932
QY 263 -----G-----Q-----HK--G-----P-----I-- 269
Db 3933 IVENEHGKDTAQCNVTVDVPGTPPEGLKIDBIHKEGCTLNWKPPDNGGTDVLHYIVEK 3992
QY 270 -----G-----F-----A-----L----- 272
Db 3993 MDTSRGTWOEVGTFPDTAKVNKLVPGEYAFRKA VNLQGESKPLEABEPIIAKNQFDV 4052
QY 273 -----KWN-----KWN-----KK-----G----- 278
Db 4053 PPVDKPEVTDWKDRIDIKWNTANNGAPVTGYIVEKKEGSAIWEAGKTPGTTFSA 4112
QY 279 -N-----F-----IL--S-----AG----- 285
Db 4113 DNLPKGEVEFRVIAVNAAGSPSPDPTDQITKARYLKPILITASRKIKAGFTHNLE 4172
QY 286 -----VD--K--TT----- 290
Db 4173 VDFIGAPDPTATWVGDSGAALPELLVDAKSTTSIFFPSAKRADSGNYKLVKNELGE 4232
QY 291 -----I-----I-----W-----D----- 294
Db 4233 DEAFIEVIQDRPSAPEGPLEVSDVTKDCSVLWKPDKDGGABISYVVEKRDTKNTW 4292
QY 295 -----A-----H-----T--G-----E-----AKQOF----- 304
Db 4293 VPUSAFVTGTSITVPKLTGEGHEYFRVMAENTFGRSDSLNLTDEPVLAKDPFGTTPGKGRP 4352
QY 305 -----P-----P-----P----- 306
Db 4353 EIVDTDNHDHIDIKWDPDRNGSGSPVDHYDIERKDAKTGRWKVNTSPVQGTAFSDTRVOK 4412
QY 307 -H-----SA-----P-----A-----LD-----V----- 314
Db 4413 GHTYFRVAVNAKAGPQSPSDSAAATAPKMPHEAPKFDLDLQKFRVKAPELVTITPF 4472
QY 315 -----D--W-----Q-----S----- 318
Db 4473 TASPOQDISWTEGGKPLACVETDTSQTKLVIPSTRSDSGPVYKIKAVNPYGEAEANIKI 4532
QY 319 -----N-----N----- 320
Db 4533 TVIDKPGAPENITYPAVSRHTCTLNWDAPKDDGGABIAKYIEQVSGQIMDKVPLGIS 4592
QY 321 -----T-----F-----A-----S--C-----ST----- 327
Db 4593 GTAYTVRGLEHQOQYFRIRAEANAVGLSDYCOQVPVVIKDPFPPGAPSTPEITGYDTNQ 4652
QY 328 -----D-----M-----CI----- 331

Db 4653 VSLAMNPPRDDGSGSPILGYVVERFEKRGGDWAPVMPVKGTECIVPGLHENETQFRV 4712
QY 332 -----H-----V--C----- 334
Db 4713 RAVNAAGHEPNSBSPVTCRPPYVEKGPADAPRVGKITKNSAELTWNRLRDLRGAPIDG 4772
QY 335 -----KLK----- 337
Db 4773 YIVEKKLGDNDWTRCNDKPVDRDTAFEVKNLGEKEEYFRVIAVNSAGEBEPKPSDLVL 4832
QY 338 -----Q-----D-----R-----PI----- 342
Db 4833 IBEQGRPIFDINNLKIDITVRAGETIQIRIPYAGGNPKPIIDLFGNSPIFENERTVVDV 4892
QY 343 -----K--T-----F-----Q----- 346
Db 4893 NPGEIVITTTGSKRSRSDAGPYKISATNKYKDTCKLVFLDAPGKTGPIRATDIOADAM 4952
QY 347 -----GH-----T-----N--B--V--N-- 353
Db 4953 TLSWRPDKNGGDAITNYVVEKRTPGGDWTVVGHVPGTTLRVNLDANTPYEYFRVRAEQ 5012
QY 354 -----AI--K-----W-----D--P----- 359
Db 5013 YGVGEPLETDDAIVAKNPPDTPGAPQPEAVETSEBAITLQWTRPTSDGGAPIOGVYIEK 5072
QY 360 -----T-----GN----- 362
Db 5073 REVGSTWTKAAFGNILDTKHRTGLTPKTYEFRVAAYNAAGQGSYSVNSVPITADNAP 5132
QY 363 -----L----- 363
Db 5133 TRPKINMGMLTRDILAYAGERAKILVPPAASPAPKVTFSKGENKISFTDPRVKVEYSDFL 5192
QY 364 -----L-----L-----AS-----C----- 367
Db 5193 ATLTIKSELTDGGLYFVELENSQSDSASIRLVKVDKSPASPOHIRVEDIAPDCTLYWM 5252
QY 368 ---SD-----DM-----T-- 372
Db 5253 PPSDGGSPITNYIVEKLDLRHSDGKWKSVFVRNLNTVVGGLIKDNRFRFRVRAETQY 5312
QY 373 -----L-----K-----I-----W-----SMK-----Q-- 380
Db 5313 GVSEPCELADVVVAKYQFVFPNQPEAPTVDKDXSTWAELEWDPDRGGS-KIIGYQVQYR 5371
QY 381 D-----N-----CVHD-----L--Q-----Q----- 389
Db 5372 DTSSGRWINAKMDLSQC-HARVTGLRQNGEPEFRIIAKNAAGFSKPSPPSERCOLKSRF 5430
QY 390 -----H-----N-----K-----EI--Y--TIKW--S----- 400
Db 5431 GPPGPPIHVGAKSIGRHNHCTITWMAPLEDGGSKITGYNVEIREYSTL-WTVASDYNVRE 5489
QY 401 P--T-----GP-----G-----T-----N 407
Db 5490 PEFTVDKLEFNDYFRVVAINAAGKIPSLPSGPIKIOESGSRPOIVVVPEDTAQPN 5549
QY 408 -----N-----P----- 409
Db 5550 RRAVFTCEAVGRPEPTARWLNRNGRELPESSRYRFEASDGVYKFTIKEVMDIDAGEVTEV 5609
QY 410 -N-----ANL----- 413
Db 5610 SNPYGSDTATANLVQAPPVIEKDVNTILPSGDLVRLKIYFSGTAPFRHSLVLRREID 5669
QY 414 M-----L-----A--S--A-----S--F----- 420
Db 5670 MDHPTIRIVEFDHILITIPALSVREAGRYEYTVVNSDSEANTGFWLNTGLPEAQGPL 5729
QY 421 -----D--S-----TV----- 424

Db 5730 HISNIGPSTATLSWRPPVTDGSKITSYVVEKRDLSKDEWVTVTSNVKMMYIIVTGLFEN 5789
Qy 425 ----- 424
Db 5790 HEYFRVSAQNEGIGAPLVSEHPPIARLPDPPTSPNLNLEIVQGGDYVTLWSQRPLSD 5849
Qy 425 --EL--W----- 427
Db 5850 GGGRLRGYVKEQBEHDEWFRCNQNPSPNNYVNPNLIDGRKYRYRVFAVNDAGLSOLA 5909
Qy 428 --D--V--D-----R----- 431
Db 5910 ELDQTLFOAGSGGPKIVSPLSDLNNEVGRCVTFECEISGSPREYRWFKGGKELVDTS 5969
Qy 432 --G--I-----C-----I--H-- 436
Db 5970 KYTLINKGKQVLIINDLTSDDAEDYTCRATNSSGSTRANLRIKTKPRVFPPIPKYHGG 6029
Qy 437 -----TL----- 438
Db 6030 YEAQGETIELKIPYKAYPQGEARWTKDGEKIENNSKFSITDDKFAITLISNASREDYG 6089
Qy 439 ----- 438
Db 6090 EYRVVENSVDGSGTVNTVADVPEPPRPPIENILDEAVILSKPPALDGGSLVTNYT 6149
Qy 439 -----KH-Q--E--PV----- 445
Db 6150 IEKREMGGSWSPCAKSRVYTTIEGLRAGQYEFRIIAENKKGQSKPCBPFTAPVLIPGD 6209
Qy 446 --Y--S-----V----- 448
Db 6210 ERKRRGYDDEQKIVRGKTVSSNYVDYVQYVQPVVEIKHDHVDHYDHEEL 6269
Qy 449 --AF-----S-----P----- 452
Db 6270 GTGAFGVHVRTERATGNFAAKFVMTPHESDKETVRKEIQTWSVLRHPTLVNLHDAFED 6329
Qy 453 D--G-----R-----Y----- 456
Db 6330 DNEMWMIYEFWGGELFEKVADEHNKMSDEAVEYMRQVCKGLCHMHHNNYVHLDLPEN 6389
Qy 457 -----L-----A----- 458
Db 6390 IMFTKRSNELKLDLFGTLAHPKQSVKVTGTGAFAAPEVAEGKPGVGYTDMWSGVVL 6449
Qy 459 --SG-S-F--D--K-C-----V 466
Db 6450 SYLLSGLSPFGGENDDETILRNKSCDMNMDSAFSGISBDGKDFIRKLLADPNTRMTI 6509
Qy 467 H--I--W-----N----- 470
Db 6510 HQALEHFWLTPGNAPGRDSQIPSSRYTKIRDSIKTKYDAMPEPLPLGRISNYSYLRKHR 6569
Qy 471 -----TQ-----T--G-A-L-----V-H----- 478
Db 6570 PQEYSIRDAFWDRSEAQPREIVKPYGTVEGQSANFYCRVVIASSPPVVTWHKDDRELKQ 6629
Qy 479 S--Y--R--G-T-----G-----I-----FE-- 488
Db 6630 SVKTMKRYNGNDYGLTINRVKDDKGEYTVRAKNSYGTKEEIVFLNVTRHSEPLKPEPLE 6689
Qy 489 -----V-----C-----W----- 491
Db 6690 PMKAPSPRVEBEFKERRSAPFFTHLRNPLIOKNHOCKLTCSLQGNPNPTIEWKDGHP 6749
Qy 492 -----NA--AG-----D-----K-----V 498
Db 6750 VDEDRVQVSPRSGVCSLEIFNARVDDAGTVTATNDLGDVDSVCVLTVQTKGGEPIPV 6809
Qy 499 -----G-----A-----SA----- 502
Db 6810 SSFPRRAYDTLSTGTDVERSHYADMRRSLRDSVPDRSAADLTKITNELPSFTA 6869

Qy 503 --SD-----GS-----V----- 507
Db 6870 QLSDSSETVEVGSAAEFSAVSGQPELIEWLHNGERISESDSRFRASYVAGKATLRISDAK 6929
Qy 508 -----C-----VL--D-----LR--K 514
Db 6930 KSDEGQVLCRANSAGQEQTRATLTGVGDQPLLNHGAGQAVESELRTVK 6978

RESULT 13
T12117

polyprotein - fava bean dsRNA replicon

C:Species: Vicia faba (fava bean)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C:Accession: T12117

R:Pfeiffer, P.

J. Gen. Virol. 79, 2349-2358, 1998

A:Title: Nucleotide sequence, Genetic organization and expression strategy of the double-

A:Reference number: Z17424, MUID:98451319; PMID:9780039

A:Accession: T12117

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-5825 <PFE>

A:Cross-references: UNIPROT:O82731; EMBL:AJ000929; NID:G3184155; PIDN:CAA04392.1; PID:G31

A:Experimental source: virion; cultivar 447

C:Comment: This gene product may be cleaved into several proteins including helicase and

C:Genetics:

A:Genome: dsRNA replicon

C:Superfamily: fava bean dsRNA replicon polyprotein

Query Match 70.3%; Score 2571.1; DB 2; Length 5825;

Best Local Similarity 9.1%; Pred. No. 3.9e-52;

Matches 442; Conservative 56; Mismatches 13; Indels 4349; Gaps 376;

Qy 1 M---SIS-----S-----D-----E-----V-----N-- 9
Db 949 MATRPSITQDAPVMKESNTLLIDMNLIRMDPMWLITECLRRKITTAYVTPMLELDQNP 1008
Qy 10 -----F-----LV-----Y----- 13
Db 1009 QTILGEVQFRKVENQIYVITAWSDRPIRVTTENYMNWTKVDAFVIGENTYPTIDTVREAN 1068
Qy 14 ---R--- ---SG--- ---PS--- ---HS-AF--- 26
Db 1069 GISLRIDLAASSTFEMSEVETVGSISISVPTVYVYKPKRKLMSGEKKVKULHREL 1128
Qy 19 -----SG----- ---FS----- ---L-QE- 18
Db 1129 YRRMMIRNVSGTSFKELVDYIGISNMKNVINDMVVSFSDISDDTIRHTAAFPVYSKMIT 1188
Qy 27 -----T--F----- --- 28
Db 1189 RANSMLTERVLESVNTARGFLASLLISOAETMLKITGLDETMLTTVNRWMSKGSVMRL 1248
Qy 29 -----G----- ---IK----- 31
Db 1249 KGMMLNLSLIDNLNLCYSREPVIHRLRAQTEILEPTGTCPHKTPDFVKECSMCECG 1308
Qy 32 -----S----- ---H----- 33
Db 1309 VNPATGAIGFCSCSLPTDKHQCDHPCKHAHEAGTRCSCLLPIAGEACPGVNRQIE 1368
Qy 34 --I-----S-----Q----- ---SNIN--- 40
Db 1369 SEILFENSESEADEABEQNRVKNRKNRPRPRDKNGRNNRNNRNDRRRTDANHHNNVYHH 1428
Qy 41 -----CALV--PP--AAL-----I-----S-----I 52
Db 1429 GHNKQRQGA--TQQPPKSAALHPHPPDDNDNDTDPSTFTVLTADPTNPPGQAPEGDNI 1487
Qy 53 -----I-----Q-----KG-----L-Q-----Y-----V 60

Db 1488 PHENDIPSPSTTQSSPPDDTNYSGPEHNRNIQNHRIIAQTNBFFGDLVAYGNDIAV 1547
QY 61 -----E-----A-----EV-----S-----I--N 67
Db 1548 KKNSHKLPNEORLBSRQFMMGDSIYATENLEVIKTYNTSGEGYCGYNALKILYPN 1607
QY 68 -----E-----D-----
Db 1608 LDLTLEMQEIVGSETOFQDWEIMRVAQAQNLNLIIVTERCALVNKSVCSNEBVGICHCR 1667
QY 70 --G-----TL-----F-----D-----G-----R-----P----- 77
Db 1668 HRGVMLEHWEAALAIQKGFADYHPTFTNALTRDLDLDFAKSSGLANKNKRVSFVLGDPRL 1727
QY 78 -I-----ESLS-LI-D--A--VM-----P-----D-----D----- 90
Db 1728 KLOTENHESLSKVDVNPAGFKIYKGTSHYVNDNRNLGYSFQYGTFTAQITEDNQEL 1787
QY 91 -----VVOT-----R-----Q-----QA----- 98
Db 1788 VELLILAYSAPARVIHTDWFDPRESVPADATOGIHDYQQLVRDIAEVNQACAGDLIKS 1847
QY 99 -Y-----R-D-K-----L-AQ-----Q-----QA----- 107
Db 1848 ELIQNRVDCKEYINNHLFKVQKTKGLKPDGLISGLVGIQSTSYVDVSHVIGYVIRTQ 1907
QY 108 -----A-A-----A-----A-----A-A-A-A----- 114
Db 1908 WCTCVTEKCAKATYIKVDLYRKNKTSGLRALFOI FRWDFRDIKTLEKATAVDAIAGWG 1967
QY 115 -----A-----AAAS-----QQG-----S----- 123
Db 1968 KSTEIVKLVNODCTVVAQTSAAVSNILEKLEQGGKMKVMKMSIEKCMTOQVNTPTLVLD 2027
QY 124 -A-----KN-----G-----E----- 128
Db 2028 EASMITWETLSLTGPQVENVLYGNTLOICVLDMYRTGSRATKSILEQAGIIRRYHTT 2087
QY 129 -----N-----T-A--N-----G-----E----- 135
Db 2088 HRIGNPLARELSLVTKELTNAKHETNCTKSWDAVRWAEKLSIAGSLEEPVILCFYNNA 2147
QY 136 -----N-----G-A-----H-----T----- 140
Db 2148 VRAVMNVLKVGCRVDTIHFKQGLEADNVILQWCPTGOTPGRITLDKHQCLSATRAEKN 2207
QY 141 -I-----ANN-----HT-----DM-----M----- 149
Db 2208 LVWISVNEYSNNVPLHKGATIGGSKHTQPDTENNLADKTLQVWSHLLSKVMIYSKDKT 2267
QY 150 -----E-----VD-----Q----- 153
Db 2268 IERLECSRLQSEEGNLVMLKNTLTHLDVDYVSGNDQHSRLRQSGKLEKFKMSLSLFS 2327
QY 154 -----D-V-----EI----- 157
Db 2328 KKPAPDHVSSSTLGKLAIEVDQANANKDRINSQIMNIYTVQKNLAIISNGSIAAAVTPVL 2387
QY 158 -----P-----P-----N-K-----AVV-----LR----- 166
Db 2388 TAMLAQKPGITRAECSEFDEPTETLTANILMLGVNVLKIQLDFRNDKVILGSGGLRSFVAD 2447
QY 167 -----CH-----E-S-----E-----V-P-IC----- 175
Db 2448 SSEIKOLYNQKFCCHYVLKYRNFTEISVKKMNFERVLMIYVFGVCLSGSLHSIV 2507
QY 176 -----AW-----N-----P-----V-----SD-----LLA 185
Db 2508 TIENELYKQSYGCSLGGGWIFHINKPIIVGDPSTYNTGNSRTVRFNSKQSDKCPILA 2567
QY 186 -----SGS-----GDS--TA-----RI-----W----- 196
Db 2568 YLRLMNVDSGTVKLPDPVGRGDFDMPFEGDNVMAALVIERVLKGPMMMAAKLKNMLTN 2627

QY 197 -----N--L--S-----E-----NS-----T-----SG-S--TOLV- 210
Db 2628 TNQDLNCFYDNNSPELFDLSLKARLESNGFTNFNTLVNVEFVGTRLEFNGFTNFNTLVN 2687
QY 211 -----L--RH--C-----I----- 215
Db 2688 VNSMDLRCRPILCTLDKRYLAALMDQDNLIVWTQISGKPIHTPAGVTRVSNKAFKMAVR 2747
QY 216 -----R-E-----G-----G-----Q-----D----- 221
Db 2748 QMLNRLFKLTNTQLSRDKPMGAPNRPGLLPKMAQHEADNDKVAVAARKYKQTOAAMM 2807
QY 222 -----V-----P-SNK-----D--V-----T 229
Db 2808 KLKSRMYLTAMWVNGNSEYASINMOQHNVTITSTNYPENNLGLVDLYVAQAIHNTGYT 2867
QY 230 S-----L--DW----- 233
Db 2868 SALYITNNACTAVLCGHWDWFCPPDDGWSTSRFLSSTHDTITNLADIKSILDSEIAKLD 2927
QY 234 -----N-----SE-----CT----- 238
Db 2928 KOEELSEETAVLKEKKERSEIRHNOILRTEDSWSESQYONVKRGTVIVSVNSCGISTES 2987
QY 239 -----L--L-----A-----TGS-----Y-- 245
Db 2988 IEKIMETTCAMEFLMCVPTLTTPSDNTTAHRVGVTGSNNVKITYPGARNTLTPNPEIVCM 3047
QY 246 -----DGF-----A-----R----- 250
Db 3048 KSGKRATVQSKKD--FDMISLHTTTVAMIASVQILKVITITRPFNEHAMPKALPIFPSPFS 3106
QY 251 -----I--W-----T--KDG-----NL-----A 259
Db 3107 HOSDIEKVVQYPRFTCNFGDGSNSLIITKRRKLVSKALLAKMNERALRHDSNLKDLQA 3166
QY 260 -----S-----T--L--G-----O--H----- 265
Db 3167 YGRSYLMTQVYTERYVNVINKNGLDEAVCAALTMKHRINEQLEHTMGLMETFRYVDT 3226
QY 266 -----K-G-P-----I-----F----- 270
Db 3227 DSITDIWVKCKNLEKVMSTANTMQEKLGPASFTGLLDEIANTQRVVFIDERTNQR 3286
QY 271 -----A--LKW-N-----KK----- 277
Db 3287 VMSWDSNIKHMSDEBSIKKKWASTADNLDWMNIGMIOBYRKVLCRKEDIPKTKRACPGF 3346
QY 278 -----GN----- 279
Db 3347 QOVVDCAKAVAOKEARKTDPGARVLVLEPIPLIGNVDITRAATOGTVTKNENVIRSE 3406
QY 280 --F-----I-----LSA-----GVD-K-----T--TII----- 292
Db 3407 ITFRARQSIRRAQTSRLRSYLDNWNEMWKAYSALBAG-DLKVMVRELSRIAIIKSTSTP 3465
QY 293 -W-----D-----A-H--T--GEA-----KO-----O--F----- 304
Db 3466 TWHLKEDLEAMLDPNLTRAHELEITKLIGDAVWSRWEQSKAEKKQNEYDFIKSDLGLST 3525
QY 305 -----P----- 305
Db 3526 LNDHDLNIVLSLLAVDRPSKVKKDLTLNTLGSQMSVNEAVMLRRKTRQWCESASNGQL 3585
QY 306 -----FH-----S-----A----- 309
Db 3586 VDHHDHANSVSTTEQNGQKLNILSLHBEAGICDINLSDASPLKTRASSVLPALITTCVQ 3645
QY 310 -----P-----ALDVD--W-----QS----- 318
Db 3646 TPGMVELTSPKVEIIEAMDNDCEVNRCTIEKYVTQNIIEPHFRITNGLRAIMOQSKMLTES 3705

QY 319 -----N-----T-----F-----AS-C-----STD--- 328
 Db 3706 QAIIVCQLLGLNCCLIQNGETGTVNFAPNPFVQLMRLSODASYDHCVLINLIGADGVK 3765
 QY 329 -----M-----C-----I-----HV-----C----- 334
 Db 3766 RLSPENIAKENWEQLEHVCVDNKPPIVAVGENPYACVSHVHIDNEDLARLTSCSQPYEKI 3825
 QY 335 -----K----- 335
 Db 3826 PVLDRFEMLPVNNVMMARMTKTSATLLKKVTPITTGAGSALATRNYYKATPMVTTPKG 3885
 QY 336 -----L-----G-Q----- 338
 Db 3886 DVLVLRDHCEKKVWCHTTITVDTPGGVEQILEITDDIGDGAVALDLGLHCLTTPADRTPR 3945
 QY 339 ---DRP---IKT---F---OG---H---T--- 349
 Db 3946 RDRPVNTSNEVWSLNIETQRYLEATKGLRSHVGQEGKLIHYHFDNREHMYDEHYIDA 4005
 QY 350 ---N---E---VN---A---IK-W---D-P--- 359
 Db 4006 MLDENIVIELDPTVOVPTGVNIVNLICIRRMFLAMLYGPTWVAYCCDNFSAQLAWFL 4065
 QY 360 -----A---S---CS---DD-M--- 371
 Db 4186 DIKQCAVYHPTGKVALRVVESHSEIACSTKVYGVDPKDDENRAGASNLTSVPASDX 4245
 QY 372 -----T---G---N---L--- 375
 Db 4246 ADPGSSQLSCEENWSQTGPVACVHIEPCDHAANASNDVLKEITQPEQNHHIISKI 4305
 QY 376 -----W---S---M--- 378
 Db 4306 LATAQAKASNPEDKSMKSGATSADLOAIQSLCDSPGMSQVADQAKLEVMNFSDEVT 4365
 QY 379 -----K-Q-----D----- 381
 Db 4366 LSGIHRORNVNSQGYLQVRDISERWAIYKGIHGHSDVPTNPNPGWMLAPPPMRKLG 4425
 QY 382 -----NC-----V-H-----D-L-----Q----- 388
 Db 4426 VYKEKEIKVTMNYAGNCAYELNRIKREPKSVFTHNVKPEVMDLILTYTSANNKQYPTN 4485
 QY 389 -Q-----H-----N-----K-E-I-----Y----- 395
 Db 4486 LORLADKYLHGRVSHVYVYEGYNRDLQTCVELTKMERVIVVTDAPDVYNQLDADIQV 4545
 QY 396 ---TIK---W---SP---TG---P--- 404
 Db 4546 REKIITIKVTSQPEPYRAIWRFWILPANIDPECELYILNGREVAENDMELVDRVMEESP 4605
 QY 405 -----G-----T-----N-----N----- 408
 Db 4606 EDFDIVSALHCOAFNVCCGYAKYNHAKFTDYNTFNDIYEAGNSYGGDEIYMSRYKVXKM 4665
 QY 409 -----P-NA-----NML-----A-----SASF----- 420
 Db 4666 WLYYIPGGAWYGIEGKWLNALDITICLKEGERTRIDHVKGLFCSANFFHVPDITKG 4725
 QY 421 ---D-STV---V-----R-----L--- 426
 Db 4726 IYDYSTVDSVDGIPWQYAMQIINGHSVTENPVHSTGESTGNPRISNDGKXPWQLNPN 4785
 QY 427 -W-D-----V-----D-----R-G-----I-----C----- 434

Db 4786 NWLLDPGLKKSEKMKIQKFLDSSSVKCIKESOSLPLOGLPKPIVVVFANQACEAIGVM 4845
 QY 435 -----I-----H-TL----- 438
 Db 4846 NNTGTLVVEVNIQAQKLAARNEPLOHOTLLIKSHLILMSTTGRHASKWPEYTHLMKADLQ 4905
 QY 439 ---T---K---H-Q---E---P---V--- 445
 Db 4906 NCQVYVEGSATSILOWKREETOHAQVYTTTEERDVFRIPQALDHSVGTLLKTVKVTGNE 4965
 QY 446 --YS--V---A---F---S---P--- 452
 Db 4966 LGLSSSSGVLARVLKNDRLAMNHNLETOHLQKLEFDRRIISHSHESSGNTPKPRIGOT 5025
 QY 453 ---D--- 455
 Db 5026 LYLADVLTAAGIECVNHITTSREWEKALKNPITKVIMPLTNSVLVSGKLGCVQPTIPR 5085
 QY 456 -Y-----L----- 459
 Db 5086 CYAMLCQGVKLKIFKPETRGGGTPSNVSANVSPNPTTAGPGDTPPLITSNWKNVESGAS 5145
 QY 460 ---G-S---FD---K--- 464
 Db 5146 NQSPFGVSAKAYQDKLREKKIETPDQLRLVLETMFLDFFDAYALMSHGLTDLRLKNDGNGWH 5205
 QY 465 C---V---H---I---W--- 469
 Db 5206 CATDATMKSGDFEIIPEVGGDLPLSLNTSALRHNETNARVIDLWDDTDLRLDWLTLYAPKNP 5265
 QY 470 ---N---TO---TG--- 474
 Db 5266 MKLTSRVSPQGGKINLKTLLTNRPCQTPVPQVQMGNGENAVTGLRGLSVLPLRREPNNVT 5325
 QY 475 -----AL----- 476
 Db 5326 HELHKERTAYRDGWERVLKDFKANTITISDADVKTWLSRSDWKALATSTIKWLTGLP 5385
 QY 477 ---VH----- 478
 Db 5386 SNPMAVNVHVHTESLLKANPIMYWRQTQGRIVMQPKELCALMSPAFIAIKRRLKEVLR 5445
 QY 479 ---S---Y----- 480
 Db 5446 DEIVYTDGLTPDMLSARATIQVYVYFEDDLVIQDRQTQOELIDLEFQWMLDLGLDLINLA 5505
 QY 481 ---R---G---TG--- 484
 Db 5506 NLWRLVHNKWRPKGQHSWGQOLDAMRLTGOATTALGNATNLVHSSSFVIEHROAIKLMFV 5565
 QY 485 -G-I-F---E---V---C----- 490
 Db 5566 LGDDNITFMSABPNLTYYKRLMSERYNNMRSPQVSRNVGTFCSLLCYRNSFGHCEVGPDP 5625
 QY 491 -----W-N- 492
 Db 5626 VRLRHRFEVTGGNHYELPVAVKMPHNVRIGKTFPNGRYLVHMENEEAEEMFVWLRNH 5685
 QY 493 A---A---G---D-K---VGAS---A--- 502
 Db 5686 ANRSYILVLKTSNAEWSRLDGVQVHDLSDKIDTSGNLAEALGIGASMGMPNPMMAAGMSY 5745
 QY 503 ---S---D---G---S---VCV---L-D---L---RK 514
 Db 5746 LHMGLKSEADKINTEKEFGLRLRWYDVASCKNATCIKYNMSPMELEDNIGELLKMRK 5805

RESULT 14

S25111

alpha-2-macroglobulin receptor precursor - mouse

N; Alternate names: CD91; iDL receptor-related protein 1; low density lipoprotein receptor
 C; Species: Mus musculus (house mouse)

C>Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 09-Jul-2004
C/Accession: S25111; S32554
R/van Leuven, F.
submitted to the EMBL Data Library, July 1992
A/Reference number: S25111
A/Accession: S25111
A/Molecule type: mRNA
A/Residues: 1-4545 <VAN1>
A/Cross-references: UNIPROT:Q61291; EMBL:X67469; NID:q49941; PIDN:CAA47817.1; PID:q49942
R/van Leuven, F.; Stas, L.; Raymakers, L.; Overbergh, L.; de Strooper, B.; Hilliker, C.;
Biochim. Biophys. Acta 1173, 71-74, 1993
A/Title: Molecular cloning and sequencing of the murine alpha-2-macroglobulin receptor c
A/Reference number: S32554; MUID:93250049; PMID:8485155
A/Accession: S32554
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-28,4416-4453 <VAN2>
A/Cross-references: EMBL:X67469
C/Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated
C/Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
C/Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glycopro
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-3944/Domain: Product: alpha-2-macroglobulin receptor #status predicted <MAT>
F:20-3944/Domain: alpha-2-macroglobulin receptor S15K chain #status predicted <S15K>
F:28-65/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:73-109/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:116-149/Domain: EGF homology <EG1>
F:155-189/Domain: EGF homology <EG2>
F:199-240/Domain: LDL receptor YWTD-containing repeat homology <YW01>
F:241-282/Domain: LDL receptor YWTD-containing repeat homology <YW02>
F:293-335/Domain: LDL receptor YWTD-containing repeat homology <YW03>
F:336-379/Domain: LDL receptor YWTD-containing repeat homology <YW04>
F:380-421/Domain: LDL receptor YWTD-containing repeat homology <YW05>
F:422-469/Domain: LDL receptor YWTD-containing repeat homology <YW06>
F:479-520/Domain: EGF homology <EG3>
F:572-614/Domain: LDL receptor YWTD-containing repeat homology <YW07>
F:615-660/Domain: LDL receptor YWTD-containing repeat homology <YW08>
F:661-711/Domain: LDL receptor YWTD-containing repeat homology <YW09>
F:712-753/Domain: LDL receptor YWTD-containing repeat homology <YW10>
F:754-800/Domain: LDL receptor YWTD-containing repeat homology <YW11>
F:808-843/Domain: EGF homology <EG4>
F:855-891/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:896-932/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:937-972/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:977-1012/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:1016-1052/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:1063-1098/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F:1103-1141/Domain: LDL receptor ligand-binding repeat homology <LDL9>
F:1146-1183/Domain: LDL receptor ligand-binding repeat homology <LDLA>
F:1186-1222/Domain: EGF homology <EG5>
F:1228-1262/Domain: EGF homology <EG6>
F:1270-1309/Domain: LDL receptor YWTD-containing repeat homology <YW12>
F:1310-1356/Domain: LDL receptor YWTD-containing repeat homology <YW13>
F:1357-1399/Domain: LDL receptor YWTD-containing repeat homology <YW14>
F:1400-1446/Domain: LDL receptor YWTD-containing repeat homology <YW15>
F:1447-1489/Domain: LDL receptor YWTD-containing repeat homology <YW16>
F:1490-1532/Domain: LDL receptor YWTD-containing repeat homology <YW17>
F:1541-1579/Domain: EGF homology <EG7>
F:1584-1627/Domain: LDL receptor YWTD-containing repeat homology <YW18>
F:1628-1670/Domain: LDL receptor YWTD-containing repeat homology <YW19>
F:1671-1714/Domain: LDL receptor YWTD-containing repeat homology <YW20>
F:1715-1754/Domain: LDL receptor YWTD-containing repeat homology <YW21>
F:1755-1797/Domain: LDL receptor YWTD-containing repeat homology <YW22>
F:1798-1847/Domain: LDL receptor YWTD-containing repeat homology <YW23>
F:1851-1887/Domain: EGF homology <EG8>
F:1935-1977/Domain: LDL receptor YWTD-containing repeat homology <YW24>
F:1978-2020/Domain: LDL receptor YWTD-containing repeat homology <YW25>
F:2021-2064/Domain: LDL receptor YWTD-containing repeat homology <YW26>
F:2065-2106/Domain: LDL receptor YWTD-containing repeat homology <YW27>
F:2107-2152/Domain: LDL receptor YWTD-containing repeat homology <YW28>
F:2160-2195/Domain: EGF homology <EG9>
F:2200-2242/Domain: LDL receptor YWTD-containing repeat homology <YW29>

F:2254-2295/Domain: LDL receptor YWTD-containing repeat homology <YW30>
F:2345-2389/Domain: LDL receptor YWTD-containing repeat homology <YW31>
F:2390-2430/Domain: LDL receptor YWTD-containing repeat homology <YW32>
F:2431-2474/Domain: LDL receptor YWTD-containing repeat homology <YW33>
F:2483-2518/Domain: EGF homology <EG10>
F:2525-2562/Domain: LDL receptor ligand-binding repeat homology <LDLB>
F:2567-2601/Domain: LDL receptor ligand-binding repeat homology <LDLC>
F:2606-2640/Domain: LDL receptor ligand-binding repeat homology <LDLD>
F:2653-2689/Domain: LDL receptor ligand-binding repeat homology <LDLE>
F:2697-2731/Domain: LDL receptor ligand-binding repeat homology <LDLF>
F:2735-2770/Domain: LDL receptor ligand-binding repeat homology <LDLG>
F:2775-2813/Domain: LDL receptor ligand-binding repeat homology <LDLH>
F:2819-2854/Domain: LDL receptor ligand-binding repeat homology <LDLI>
F:2859-2898/Domain: LDL receptor ligand-binding repeat homology <LDLJ>
F:2905-2940/Domain: LDL receptor ligand-binding repeat homology <LDLK>
F:2945-2981/Domain: EGF homology <EG11>
F:2987-3022/Domain: EGF homology <EG12>
F:3030-3069/Domain: LDL receptor YWTD-containing repeat homology <YW34>
F:3070-3114/Domain: LDL receptor YWTD-containing repeat homology <YW35>
F:3115-3157/Domain: LDL receptor YWTD-containing repeat homology <YW36>
F:3158-3201/Domain: LDL receptor YWTD-containing repeat homology <YW37>
F:3202-3242/Domain: LDL receptor YWTD-containing repeat homology <YW38>
F:3243-3285/Domain: LDL receptor YWTD-containing repeat homology <YW39>
F:3295-3331/Domain: EGF homology <EG13>
F:3335-3370/Domain: LDL receptor ligand-binding repeat homology <LDLL>
F:3375-3409/Domain: LDL receptor ligand-binding repeat homology <LDLM>
F:3414-3449/Domain: LDL receptor ligand-binding repeat homology <LDLN>
F:3454-3490/Domain: LDL receptor ligand-binding repeat homology <LDLO>
F:3493-3532/Domain: LDL receptor ligand-binding repeat homology <LDLP>
F:3537-3571/Domain: LDL receptor ligand-binding repeat homology <LDLQ>
F:3576-3610/Domain: LDL receptor ligand-binding repeat homology <LDLR>
F:3614-3648/Domain: LDL receptor ligand-binding repeat homology <LDLS>
F:3655-3691/Domain: LDL receptor ligand-binding repeat homology <LDLT>
F:3696-3732/Domain: LDL receptor ligand-binding repeat homology <LDLU>
F:3742-3777/Domain: LDL receptor ligand-binding repeat homology <LDLV>
F:3786-3823/Domain: EGF homology <EG14>
F:3829-3861/Domain: EGF homology <EG15>
F:3869-3912/Domain: LDL receptor YWTD-containing repeat homology <YW40>
F:3913-3970/Domain: LDL receptor YWTD-containing repeat homology <YW41>
F:3945-4545/Domain: alpha-2-macroglobulin receptor 85K chain #status predicted <85K>
F:3945-4545/Domain: 85K chain extracellular #status predicted <EX1>
F:3971-4013/Domain: LDL receptor YWTD-containing repeat homology <YW42>
F:4014-4057/Domain: LDL receptor YWTD-containing repeat homology <YW43>
F:4058-4100/Domain: LDL receptor YWTD-containing repeat homology <YW44>
F:4101-4143/Domain: LDL receptor YWTD-containing repeat homology <YW45>
F:4152-4183/Domain: EGF homology <EG16>
F:4201-4232/Domain: EGF homology <EG17>
F:4237-4268/Domain: EGF homology <EG18>
F:4273-4304/Domain: EGF homology <EG19>
F:4309-4340/Domain: EGF homology <EG20>
F:4345-4375/Domain: EGF homology <EG21>
F:4378-4409/Domain: EGF homology <EG22>
F:4422-4445/Domain: transmembrane #status predicted <TM>
F:4446-4545/Domain: intracellular #status predicted <INT>
F:167,2999/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:2959/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
F:4076,4126,4279/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.2%; Score 2569.5; DB 1; Length 4545;
Best Local Similarity 9.7%; Pred. No. 2.6e-52;
Matches 420; Conservative 69; Mismatches 17; Indels 3807; Gaps 358;

QY	1	MS	-----I-----SSDE-----V-----N-F-I-----11
DB	89	MSRLNGIQDCMWDGSDGAHCHRELNRANCRMCQHCVCPTSPGTCYCNSSFOLEADGKT	148
QY	12	-----VY-----R-----Y-LQ-----E-----18	
DB	149	CKDFDECSYGTCSQLCTNDGSGFTCCGCGVEGVLQLQPDNRCKAKNEPVDPRPVLIIANSQ	208
QY	19	-----SG-----FS-----H-----SA-----F 26	
DB	209	NILATVLSGAQVSTITPTSTRQTMDFSYANETVCVWHVGDSSAAQTQLKCARMPGLKGF	268

QY 27 ----TFGIK-S--HISQ-----S-N-----I-----NG----- 41
Db 269 VDEHTIINISLSLHVEQWADLWLTGNFYFVDDIDDRIFVQCNRGDTCTVTLDDLELYNPKG 328
QY 42 -AL-----VP-----P-----A-A- 48
Db 329 IALDPAMKVFYTDYQIPKVERCDMGQNRKLVDSKIVPFGHITLIDLVSRLVYWADAY 388
QY 49 -LI-----SIIQK-L-Q--Y--V-B-----A-----EVS-I-----N 67
Db 389 LDYIEVVDYEGKGRQTIQ--GILIEHLYGLTVFENLYATNSDNANTQOKTSVIRVNRFN 447
QY 68 -E-----D--CTL----- 72
Db 448 STEYQVTRVDKGGALHIYHQRORPRVRSACENDQYKPGGCSIDICLLANSHKARTCRC 507
QY 73 -F-----DG-----R-----PIE-----SL-- 81
Db 508 RSGFSLGSGSKCKPEHELFLVYKGRPGIIRGMDMGAKVPDEHMIPIENLNPALDF 567
QY 82 -----S-LI-----D-----AV--M-----PD-- 90
Db 568 HAETGFYFADTTSYLIQKIDGTERETILKDGIIHNVEGVAVDMGDNLYWTDDGPKKT 627
QY 91 V-V-----QTR----- 97
Db 628 ISVARLEKAAQTRKTLLEGKMTHPRAIVDPLNGMYMTDWEEDPKDSRRGLERAWMDG 687
QY 98 -----A-Y-R-----D-K-----A-QQ----- 102
Db 688 SHRDIFVTSKVLWPNGLSLDIPAGRLYWDAFYDRIETILLNGTDKIVYEGPELNHAF 747
QY 103 -----O-A-A-----L-----A-QQ----- 106
Db 748 GLCHGNYLFWTEYRSGSVYRLERGAVAGAPPTVTLRSEPPIFEIRMYDAHEQQVGTNK 807
QY 107 -----O-A-A-----A-----A-A----- 112
Db 808 CRVNGCSSLCLATPGSROCAQADQVLTDTGVTCLANPSYVPPQCGQGFACANNRC 867
QY 113 -----A-A----- 114
Db 868 IQERWKGDGNDCLNDSDEAPALCHOHTCPSDRFKCENNRCPINRWLCGDNDGCGNSE 927
QY 115 -A--A-----A-----AS-----Q----- 120
Db 928 SNATCSARTCPNPFQSCASGRCPISWTCLDDDCGDRSDESASCAYPCTFPLTQFTCNN 987
QY 121 -----A-----Q-----G-----S-- 123
Db 988 GRCININWCDNDCGDSDEAGCSHSCSSTQFKCNSGRCPFEHWTCDGNDGCGYSDE 1047
QY 124 -----A-----K----- 125
Db 1048 THANCTNOATRPPGGCHSDEFOCPDLGLCPLRWCDGDTDCDMSDESKSCGVTHVCDP 1107
QY 126 N-----G-----EN-----T--AN----- 132
Db 1108 NVKFGKDSARCISKAWVCDGSDCEDNDSDEENCEALACRPPSHPCANNTSVCLPPDKLC 1167
QY 133 -G-----E-E-----NG-----A-----HT--I----- 141
Db 1168 DGKDDCGDGSDEBELCDQCSLNGGCHNCVAPBEGIVCSCPLGMELGSDNHTCQIQSY 1227
QY 142 -A----- 143
Db 1228 CAKHLKCSQKQDKFQSVKSCYEGWVLEPDGETCRSLDPFKLFIIFSNRHEIRRIDLHK 1287
QY 144 -----N-----H-----TD-----MM----- 149
Db 1288 GDYSVLVPLGRLNTIALDFHLSQSALYWTDAVEDKIVRGKLLDNGALTSFEVVIQYGLATP 1347

QY 150 -----EV--D-----GDVE-----I--P----- 158
Db 1348 EGLAVDWIAGNIYVWESNLDQIEVAKLDGTGLRFTLLAGDIEHPRAIADPRDGLFTWDW 1407
QY 159 -P-----N-----K-----AV----- 163
Db 1408 DASLPRIEASMSGAGERTIHRGTSGGCANGLTVDYLEKRIILWIDARSDAIISARYDGS 1467
QY 164 -VLRGHE--S-----EV-----F-- 173
Db 1468 GHMEVLRGHEFLSHPPFAVTLYGGEVYTDWRTWTLAKANKWTGHNVTVQRTWTQPDFLQ 1527
QY 174 I-----C-A-----W-----N----- 178
Db 1528 VYHPSQPMAPNCPCEANGRGPCSHLCLINYNRTVSWACPHLMKLHKDNTTCYEFKKFL 1587
QY 179 -P-----VS-----D-----L-L-A-----S----- 186
Db 1588 YARQMEIRGVOLDAPYNYIISFTVPDIDNVTVLDYDAREQRYVWSDVTOAIKRAFING 1647
QY 187 -G-----S-----G--D-----S--T-----AR----- 194
Db 1648 TGVETVVSADLPNAHGLAVDWVSRNLFMTSYDTNKKQINVARLDGSKNAVVOGLEOPHG 1707
QY 195 --I-----W-----NLS-EN--S--T-----S----- 204
Db 1708 LVVHPLRGKLYWTDGDNISMANNDGSHLLTFSGQKGPVGLAIDFPESKLYWISSGNHTI 1767
QY 205 -GS-----T--QL-----C-I-----R----- 216
Db 1768 NRCNLGSELEVIDTMRSQLKATALAIMGDKLWADQVSEKMGTCNKADGSGSVLRNS 1827
QY 213 -----H-----C-I-----R----- 216
Db 1828 TTLVHMVKVDESIOLEHEGTNPCSVNNGDCSQLCLPTSETTRSCMTAGSYLSRQQAC 1887
QY 217 EG-----G-Q-----DVPSN-K-D--V--TSL----- 231
Db 1888 EGVGSFLYSVHEGIRGIPLD-P-NDKSDALVPVSGTSLAVGIDFHAENDTIYWDNGLS 1945
QY 232 -----D--W-----NS-----EG-TL--LATGS-Y--D-GF-----AR----- 250
Db 1946 TISRAKRQDTWREDVVVNGIGRVEGLAVDWIA-GNIYWTQDGFVIEVARLNGSPRYVI 2004
QY 251 -----I-----WTK-----DG-----N----- 257
Db 2005 SQGLDKPRAITVHPEKGYLFWTEWGHYPRIERSRLDGTERRVVLNVNVSISWPNGISVDYQ 2064
QY 258 -L-----A-----STLG-----Q-H-----K- 266
Db 2065 GKLYWCDARMDKIERIDLET-GENREVLSNNMDMFSVSVFEDFIYWSDRTHANGSIKR 2123
QY 267 G-----P-----I--F----- 270
Db 2124 GCKDNATDSVPLRTGIGVOLKDIKVENRDRQKGTNVCAVANGCQQLCYRGGORACAC 2183
QY 271 A--L-----KW-----N-----KK----- 277
Db 2184 AHGMLAEDGASCREYAGVLLYSERTILKSIHLSDERNLNAPVQPFEDPEHMKXVIALAFD 2243
QY 278 --G-----N-F--I--LSAG-----V--D--KTTII-----WDA----- 295
Db 2244 YRAGTSPGTPNRIFFSDIHF--GNIQQINDGSGRTTIVENVGSVEGLAYHGRWDTLYWT 2301
QY 296 -----HT-----G--E-----A--KQ-Q--F----- 304
Db 2302 SYTTSTIRHTVDQTRPGAFERETVITMSGDDHPRAFLVDELCEQNLMMFNWNLHPSIMR 2361
QY 305 -----P-----F-----H-----SAPA 311
Db 2362 AALSANVLTLEKDIRTPNGLAIDHRAEKLYFSDATLKDIERCEYDGSRRVILKSEPV 2421
QY 312 ---LDV-----DW-----Q-----S 318

A;Cross-references: EMBL:U24703; NID:G902486; PID:G902487
 R.D'Arcangelo, G. EMBL Data Library, April 1995
 submitted to the EMBL Data Library, April 1995
 A;Reference number: S71844
 A;Accession: S71844
 A;Molecule type: mRNA
 A;Residues: 1-215; T, 217-1905; S, 1907-3355; V, 3357-3391, N, 3393-3461 <DA2>
 A;Cross-references: EMBL:U24703; NID:G902486; PID:G902487
 F;1-27/Domain: signal sequence #status predicted <Sig>
 F;28-3461/Product: reelin #status predicted <Mat>
 F;1769-1795/Domain: EGF homology <EGF>

Query Match 70.2%; Score 2568.8; DB 2; Length 3461;
 Best Local Similarity 11.8%; Pred. No. 1.5e-52;
 Matches 393; Conservative 92; Mismatches 16; Indels 2842; Gaps 361;

Qy	1	M-----SI-----S-----S-----D-----EV-----8	
Db	1	MERCWAPRALVLAALLATLARAATGYVPFSPFFLCTHGELEGDEQGEVLISL 60	
Qy	9	N-----F-----LV--YRY-----LQ--ES-----GF--S-H 23	
Db	61	HIAGNPTYYVPGQBYHVTISTSTFFDGLLVTLG-LYTSTSIQSSQISGSSAFGFGIMSDH 119	
Qy	24	S-A-----F-----T-----F-----G--I-K-----31	
Db	120	QFGNQFMCVVASHVSHLPTNLSFVWIAPAGTCVNFATATHRQGVIFKDALAQOLC 179	
Qy	32	SH-----I-----S-Q-----SNI-----N-----GAL-----43	
Db	180	EQGAPTEATAYSHLAEIHSDDFSDYSQLELNPWVECSNCEMGECQCGTIMGH 239	
Qy	44	V--P--P-----A-AL--I-----ST-----IQ--54	
Db	240	AVTECEPYGPRELTTCLNTTASVLQFSIGSGSCRSFSDPSTVSYAKNNTADMIQLE 299	
Qy	55	KG-----LQY-----V-EA-E-----VSN-----67	
Db	300	KIRAPSNVSTVIHLYLPEEAKGESVQFMQKDSLRVGEVVEACWALDNILV-INSARE 358	
Qy	68	ED-----GT-LP-----DG-----R-----76	
Db	359	VLENDLDPVDTGNWLFPGATVKHSCQSDGNSIYFHNGESBFNFATTRDVLSTEDIQ 418	
Qy	77	P-----I-----ES-LSLI-----DA-----86	
Db	419	EQWSEEFESQPTGWDILGAVVGADCGTVESGLSLVFLKQGERKLCPTPYMDTGYGNLRFY 478	
Qy	87	VM-----P-----D-----V-----QT-----94	
Db	479	FVMGGICDPGVSHENDIILYAKIEGRKEHIALDTLTYSYKVPSPVSVVINPELQTPATK 538	
Qy	95	Q-----QAV-R-----D-----K-LAQ-----Q-----106	
Db	539	FCURKSHQGNRNWAVDFHVLVPLPTSMHMIQFSINLGCCTHQPQNSVSLEFSTNH 598	
Qy	107	QA-----A-----AAA-----AA-----114	
Db	599	GRWSLLHTECLPEICAGPHLPSTVYSSENYSGWNRITIPNALTRDRTRWRQTGP 658	
Qy	115	A-----A-----A-AASQ-----Q--GS 123	
Db	659	ILGNMWAIDNVYIGPSCLKFCSGRGQCTRHCKCDPFGSPACEMASQTPPMEISESPGS 718	
Qy	124	AK-----N--G-E-----N-----129	
Db	719	ARLSSVHNFYIRGAEVSCFVGKALVFNKGRRLQITSLDSSQSRFLQFTLRIG 778	
Qy	130	T--A-----NG--E-E-E-----N-----G--137	
Db	779	SKSVLSTCRAPDQPGEGVLLHYSVDNGITWKLLEHYSYVNYHPRISIVELPDARQFGI 838	
Qy	138	A--H-----T-----I--AN-----N-----H- 145	

Db	839	QFRWQPYTHSQGEDVWAIDEIVMTSVLFNSISLDFTNLVEVTQSLGFLGNVQPYCGHD 898	
Qy	146	T--D-----M-----MEVDG-----D--V--E-----156	
Db	899	WTLCTGCDKSLASSMYVETOSMOI-GASYMIQPSLVMGCGQKYTPHMDNOVKLEISANTH 957	
Qy	157	I-----P--PN--K-----A-----VVL-----R-----166	
Db	958	GLTWHLVQEBECLPMPSCOEFTSASIHASEFTQMRVRVTVPQKTSWGTATRFWSQSY 1017	
Qy	167	GH-----CH-----E--SE-----V--F-----173	
Db	1018	TAQDEWALDNIIYGOQCPNMCSGHSCDHGVCKCDGOYGTGTECHPEALPSTIMSDFNP 1077	
Qy	174	I-----I-----C-----A-----W-----177	
Db	1078	SSWEDWQEVIGGEVVKPEQCGVSSGSLYFSKAGKQLVSWDLDTSMVDVQFVQI 1137	
Qy	178	N--P-----V-----SD-----L-L-ASG-----187	
Db	1138	GGESAACNKPDSREEGILLQYSNNGGIQWHLAEMFYSDFSKPRFVYLEPRAAGTCTR 1197	
Qy	188	SG-----D--S-----T-----A-----R--IW- 196	
Db	1198	FRWKVPVFGEDYDQWAVDDIILSEKQKQVVPVNPITLPQNFYKPAFDYPMNQMSWL 1257	
Qy	197	N--LSN-S-----T-S--G-S-----T-----Q--L-----209	
Db	1258	MLANEGMAKNDSCATTPSAMVFGSDGRFAVTRDLTLKPGVQLQKLNIGTQSFSST 1317	
Qy	210	V-LR--H-----C-----IR--EG-----GQD-----V 222	
Db	1318	APVLQYSHDAGMFWLLKEGCFPASAKGCEGNSRELSEPTVYTYG-DFEETRITIAI 1376	
Qy	223	P-----S--N-----K-----D-----227	
Db	1377	PRSLASKTRFRWIOBSSQKNVPPGLDGVISBPSPSCYSGHGDICISGVCFDLGYTA 1436	
Qy	228	V--T--S--LD-----W-----NSBG-TL--L-----A- 241	
Db	1437	AGTCTCVSNTPNHSEMPDRPEGLKSLPLWKITGGQVGTGCTILN-DGRSLYFNLGLKREAR 1495	
Qy	242	T-----GS-----Y--D-G-----F 248	
Db	1496	TVPLDTRNLISLVQFYIQGSKTSGITYITPRARYEGLVVOYNDNGILMHLRLREDFMF 1555	
Qy	249	A--RW-----A--RW-----T-----K-DG-- 256	
Db	1556	LEPQIISIDLPREAKTPATAFR-WWQPOGHKHAQWALGDVLIGVNDSSQTGFQDKLDGS 1614	
Qy	257	N-----LAS--T-L-----G-Q-----H-----KG- 267	
Db	1615	IDLOANWRIQGVDDIDCL-SMDTALIFTENIGNPRYAETWDFHVSESSFLQWEMNMG 1673	
Qy	268	PIF--A--L-----K-W--NKK-----G-----278	
Db	1674	SKP-FSGAHGIGLOQLNNGKDWLVTECPVTIGCVHYTESSTYTSERFQNRVTVY 1732	
Qy	279	N-----P--I--LSAGVDKT-TI--I-----W-----DA-----295	
Db	1733	LPLATNSPRTRFRWITQNTYTVGAD-SWAIDNVILASGCPMCSGRGICDSGRVCVDRGPG 1791	
Qy	296	H-----H-----T-----GE-----299	
Db	1792	GFPCVVPVPLPSILKDDFNGLHPLMPEVYGAERGNLNGETIKSGTCLIFKGEGLRMLI 1851	
Qy	300	AK--Q--QF-----P-----F-----H-----SA-----309	
Db	1852	SRDLCTNTMYQFSRLFIKATKTPERSHILLQFSVSGGVTHLMDPEVFPQTTRILFIN 1911	
Qy	310	P-----A--L-----D-----V-----D-W 316	

```

Db 1912 VPLPYGAQTNRFLRWPYNNGKBEIWIIDDPIIDGNLNNPVLILLDTDFGPRDNW 1971
Qy 317 -----Q--S-N-NT----- 321
Db 1972 FFYPGGNIGLYCYPSSKGAPEEDSAMVFSVNEGEHSITTRDLSVNEHTIIQFEINVCSS 2031
Qy 322 -----F-A-----S-CST----- 327
Db 2032 TDSSADPVRLEPSRDFGATWHLPLCYHSSLSVSLCSTEHPSSTYYAGTTQWRRE 2091
Qy 328 -----D-----MC-IH-----VC 334
Db 2092 VHFGLHLGCVSRFRWYQFYFAGSQPVTAIDNVYIGQCCEMCGYHSCINGTKCIC 2151
Qy 335 -----K--L--GO--DR-----P-----I-KT-----F-Q-G- 347
Db 2152 DPGYSGPTCKISTKNPDLKDDFEGGLESDRFLMSGGRPSKCGILSSGNLFFNEDGL 2211
Qy 348 -----H-----T-N-----EV----- 352
Db 2212 RMLVTRDLDLSHARFVQFMRLGCGKVPDRPSQPVLLQYSLNGLSWSLLQEFLSNSS 2271
Qy 353 NA--I-----K-----W----- 357
Db 2272 NVGRYIALEMPKARSSTLRWQPSENGHFYSPWVIDQILIGNIGNTVLEDDFSTL 2331
Qy 358 D-----P--TG-----N--LL-----ASCS--D 369
Db 2332 DSRKWLHPGCTKMPVCGTGDALVIEKASTRYVTTDIANNEDSFLOIDFAASCSTVD 2391
Qy 370 -----DM-----T-----L-KI-----W----- 376
Db 2392 SCVAIELEYSVDLGLSWHPLVRDCLPTNVECSRHYLQRIILVSDTFNKWTRITLPLSYTR 2451
Qy 377 S-----M-----KO-----DN-----CV-----H-----D-- 386
Db 2452 SQATFRWHQAPDFKQOTWAIIDNVYIGDCLDMCGHRCVQGVSCVDEQWGLYCDER 2511
Qy 387 -----L--Q-----O-----H----- 390
Db 2512 ETSPLTQKDNFRAPSNQNLTVSGKLSTVCGAVASGLALHFGSGCSRLLVTVDLNL 2571
Qy 391 NKE-I-----Y-----T-----IKW-----S-PTG-----P 404
Db 2572 NAEFIQFYMYGCLITPSNRNQVLEYSVNGGITWLLMEIFYDQYSKP-GFVNILLPP 2630
Qy 405 -----GT-----NNP-----N--A--NLMLA-SA-----SF-----D 421
Db 2631 DAKBIATFRFWMQ--PRHDGLDQNDWAIIDNVLISSADQRTVMLDTFSSAPVQHERSPAD 2689
Qy 422 S-TV-R-----L-----W--DVD--R-----GI--C-IH--TL--TKH- 441
Db 2690 AGPVGRIAFEMFLEDKTSVNNWLFHD-DCTVERFCSDPDGVMLCGSHDGREVIYAVT-HD 2747
Qy 442 -----Q-----E-P-----V-YSAF-----S--D----- 453
Db 2748 LPTENWIMQFKISVCKVPEKIAQNIHVQFSTDFGVSWSYLVPOCLPADPKCSGSVSQ 2807
Qy 454 -----G--R--Y-L-AS--G-----S-----F-----D- 463
Db 2808 PSVFFTEGWMKRITYPLPESLTGNPVRFRFYQYSDVQWAIIDNVYILPGCLDNCGGHDC 2867
Qy 464 -K--CV-----H-----I-----W-----NT- 471
Db 2868 LKEQCICDPGSGPNCYLTHSLKTLKERFDSSEIKPDLMMSLGSGSTCTECGVLAENTA 2927
Qy 472 -----Q--T-----GA--L-----V-----H-----S----- 479
Db 2928 LYFGSSTVRQAITQDLDLRGAKFLOVWGRIGSENNTSCHRVPVCRKEGVLLDFSTGGIT 2987
Qy 480 -----Y--R-----G-----T-----G--GI----- 486
Db 2988 WTLHEMDFKYISVRHVDILLPEGALTNTTLRWMPQFVINSGLVSVGVGACAVGTGQH 3047

```

```

Qy 487 -----F-----E-----V-----C-----W-----NA 493
Db 3048 SDWWSRHPSQLVDITFDDSGSSHEENWSFYFNAVRTAGCGNPSFHYLWPNKKKDKTHNA 3107
Qy 494 -----A-----GD-----K----- 497
Db 3108 LSSRELIQPGYMMQFKIIVVGCEATSCGDLHSLVMLEYTKDARSDSQVLQVOTQCLPSSNS 3167
Qy 498 VG-----A-----SA-----S--D-- 504
Db 3168 IGCSPQFHEATYINAVNSSWKRIITIQLPDHVSSATQFRWIOKGEETEKSQSWAIDHVV 3227
Qy 505 -G-----S-----VC-----LDL-R--K 514
Db 3228 IGEACPKLCSGHGYCTTGAVCICDESFOGDDCSVFSHELPSYIKDNFESARVTEANWETI 3287
Qy 509 ---V-----LDL-R--K 514
Db 3288 QGCVIGSGCGQLAPYAHGDSLYFNGCQIRQAATKPLDLTRASK 3330

```

Search completed: January 3, 2005, 15:26:59
Job time : 101.667 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:55:11 ; Search time 102.333 Seconds
(without alignments)
2889.993 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 3659

Sequence: 1 MSISDEVNFLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3650	99.8	514	1 TBLR HUMAN	Q9bk77 homo sapien
2	3618	98.9	514	1 TBLR MOUSE	Q8bhj5 mus musculus
3	3589.2	98.1	522	2 Q6GPC6	Q6gpc6 xenopus lae
4	3580.5	97.9	519	2 Q7S2M9	Q7szm9 xenopus lae
5	3418.1	93.4	527	1 TBLX MOUSE	Q9qxe7 mus musculus
6	3355.2	91.7	526	1 TBLX HUMAN	O60907 homo sapien
7	3272.4	89.4	522	1 TBLY HUMAN	O9bq87 homo sapien
8	2918.7	79.8	511	2 Q7Q371	Q7q371 anopheles g
9	2917.1	79.7	700	1 EBI DROME	Q95tj9 drosophila
10	2614.8	71.5	5636	2 Q96RW7	Q96rw7 homo sapien
11	2611.8	71.4	34350	2 Q8W242	Q8w242 homo sapien
12	2611.7	71.4	7073	2 AAP49011	Aap49011 sars coro
13	2610.5	71.3	4599	1 LR1B MOUSE	Q9j118 mus musculus
14	2607.7	71.3	20925	2 Q8NJK1	Q8njx1 trichoderma
15	2605.9	71.2	7191	2 Q6XA09	Q6xa09 alternaria
16	2605.9	71.2	7191	2 AAP87735	Aap87735 alternari
17	2602.7	71.1	7073	2 AAP82978	Aap82978 sars coro
18	2600.7	71.1	7073	1 RIAB CVHSA	P59641 h replicase
19	2600.7	71.1	7073	2 Q6JH48	Q6jh48 sars corona
20	2600.7	71.1	7073	2 Q6R7Y8	Q6r7y8 sars corona
21	2600.7	71.1	7073	2 Q6RCW7	Q6rcw7 sars corona
22	2600.7	71.1	7073	2 Q6RCX8	Q6rcx8 sars corona
23	2600.7	71.1	7073	2 Q6RCY9	Q6rcy9 sars corona
24	2600.7	71.1	7073	2 Q6RD00	Q6rd00 sars corona
25	2600.7	71.1	7073	2 Q6RD11	Q6rd11 sars corona
26	2600.7	71.1	7073	2 Q6RD22	Q6rd22 sars corona
27	2600.7	71.1	7073	2 Q6RD33	Q6rd33 sars corona
28	2600.7	71.1	7073	2 Q6RD44	Q6rd44 sars corona
29	2600.7	71.1	7073	2 Q6RD55	Q6rd55 sars corona
30	2600.7	71.1	7073	2 Q6RD66	Q6rd66 sars corona
31	2600.7	71.1	7073	2 Q6TPE9	Q6tpe9 sars corona

RESULT 1

ID	TBLR HUMAN	STANDARD;	PRT;	514 AA.
AC	Q9BZK7; Q9H211; Q9H9A1;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	01-OCT-2004 (Rel. 45, Last annotation update)			
DE	P-box-like/WD-repeat protein TBLR1 (Nuclear receptor corepressor/HDAC3 complex subunit TBLR1) (TBL1-related protein 1).			
DE	complex subunit TBLR1 (TBL1-related protein 1).			
GN	Name=TBLR1; Synonyms=IRAL;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR COMPLEX WITH NCOR1; NCOR2; GPS2; TBL1X AND HDAC3.			
RP	MEDLINE=21930350; PubMed=11931768;			
RA	Zhang J., Kalkum M., Chait B.T., Roeder R.G.;			
RT	"The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";			
RL	Mol. Cell 9:611-623 (2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	PubMed=11063877;			
RA	Zhang X., Dormady S.P., Basch R.S.;			
RT	"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors.";			
RL	Exp. Hematol. 28:1286-1296(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	PubMed=14702039; DOI=10.1038/ng1285;			
RA	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hirakawa S., Chiba Y., Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y., Togiya S., Komai F., Harara R., Takeuchi K., Arita M., Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S., Yoshioka K., Matsunawa H., Ichihara T., Shiohata H., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto Y.,			

RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu I.,
RA Mizushima-Sugano J., Sato T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maehuo Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45 (2004).
RL [4]
RN SEQUENCE OF 1-125 FROM N.A.
RC TISSUE=Adrenal cortex;
RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Marra M.A.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [5]
RP COMPONENT OF THE N-COR COMPLEX WITH TBL1X; CORO2A AND HDAC3, AND
RX HISTONE-BINDING.
RX PubMed=12628926;
RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,
RA Wong J.;
RT "Purification and functional characterization of the human N-Cor
RT complex: the roles of HDAC3, TBL1 and TBL1X.";
RL EMBO J. 22:1336-1346 (2003).
RN [6]
RP FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.
RX PubMed=14980219;
RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;
RT "A corepressor/coactivator exchange complex required for
RT transcriptional activation by nuclear receptors and other regulated
RT transcription factors.";
RL Cell 116:511-526 (2004).
CC -|- FUNCTION: F-box-like protein involved in the recruitment of the
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
CC transcription units. Plays an essential role in transcription
CC activation mediated by nuclear receptors. Probably acts as
CC integral component of the N-Cor corepressor complex that mediates
CC the recruitment of the 19S proteasome complex, leading to the
CC subsequent proteasomal degradation of N-Cor complex, thereby
CC allowing cofactor exchange, and transcription activation.
CC -|- SUBUNIT: Component of the N-Cor repressor complex, at least
CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.
CC Probable component of some E3 ubiquitin ligase complex. Interacts
CC with histones H2B and H4.
CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -|- TISSUE SPECIFICITY: Ubiquitous.
CC -|- DOMAIN: The F-box-like domain is related to the F-box domain, and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes (By similarity).
CC -|- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -|- SIMILARITY: Contains 1 F-box-like domain.
CC -|- SIMILARITY: Contains 8 WD repeats.
CC -|- SIMILARITY: Contains 1 Lish domain.
CC -|- CAUTION: Ref.4 sequence differs from that shown due to a potential
CC poly-A sequence starting in position 125.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF314544; AAK00301.1; --
CC EMBL: AF268193; AAG44736.1; --
CC EMBL: AK022956; BAB14331.1; --
CC EMBL: BC060320; AAH60320.1; ALT_TERM.
CC HSSP: P16649; 1ERJ.
CC MIM: 608628; --
CC InterPro: IPR006594; Lish.
CC InterPro: IPR011047; Quin_alc_DH_like.
CC InterPro: IPR001680; WD40.
CC Pfam: PF04000; WD40; 8.
CC PRINTS: PR00320; GPROTEINRPT.
CC ProDom: PD000018; WD40; 3.
CC SMART: SM00667; Lish; 1.
CC SMART: SM00320; WD40; 8.
CC PROSITE: PS00896; Lish; 1.
CC PROSITE: PS00678; WD_REPEATS_1; 4.
CC PROSITE: PS00892; WD_REPEATS_2; 6.
CC PROSITE: PS50294; WD_REPEATS_REGION; 1.
CC Activator; Nuclear protein; Repeat; Transcription regulation;
CC Ub1 conjugation pathway; WD repeat.
CC DOMAIN 4 36
CC F-box-like.
CC FT DOMAIN 41 86
CC REPEAT 167 206
CC REPEAT 223 262
CC REPEAT 264 303
CC REPEAT 306 344
CC REPEAT 347 386
CC REPEAT 389 437
CC REPEAT 440 479
CC REPEAT 481 513
CC DOMAIN 108 118
CC CONFLICT 31 31 E -> K (in Ref. 2).
CC CONFLICT 59 59 Y -> H (in Ref. 3).
CC CONFLICT 389 389 A -> Q (in Ref. 2).
CC SQ SEQUENCE 514 AA; 55595 MW; 0B556D2EE4BA796D CRC64;
CC
CC Query Match 99.8%; Score 3650; DB 1; Length 514;
CC Best Local Similarity 99.8%; Pred. No. 1.6e-65;
CC Matches 512; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 1 MSISDEVNPLVRYLQESGFHSFTFGIKSHISQSNINGALVPPAALISIIQGLQYV 60
CC DB 1 MSISDEVNPLVRYLQESGFHSFTFGIESHISQSNINGALVPPAALISIIQGLQYV 60
CC QY 61 EAEVINEGDTLPDGRPIESLSLIDAMPDVVQVQAVRDVKLAQQOAAAAAASQ 120
CC DB 61 EAEVINEGDTLPDGRPIESLSLIDAMPDVVQVQAVRDVKLAQQOAAAAAASQ 120
CC QY 121 QGSAKNGENTANGENGATIANNNHTDMEVDGVEIPPNKAVVLGRHSEVEFICAWNPV 180
CC DB 121 QGSAKNGENTANGENGATIANNNHTDMEVDGVEIPPNKAVVLGRHSEVEFICAWNPV 180
CC QY 181 SLLASGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGLL 240
CC DB 181 SLLASGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGLL 240
CC QY 241 ATGSDGFARITWQGNLSTLQGHKGPFPALKWNKGNFILSAGVDKTIINDAHTGEA 300
CC DB 241 ATGSDGFARITWQGNLSTLQGHKGPFPALKWNKGNFILSAGVDKTIINDAHTGEA 300
CC QY 301 KQGFPHSAPALVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
CC DB 301 KQGFPHSAPALVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360


```
OY 361 GNLASCSDDMTLKWSMKQDCNCVHDLQOHNKEIYTKSPGTGNNPNANMLASGF 420
Db 361 GNLASCSDDMTLKWSMKQDCNCVHDLQOHNKEIYTKSPGTGNNPNANMLASGF 420
OY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPSPGRVSLASGSPKCVHINTQTGALVHSY 480
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPSPGRVSLASGSPKCVHINTQTGALVHSY 480
OY 481 RTGGTGFIFVCVNAAGDKVGSASDGSVCVLDLRK 514
Db 481 RTGGTGFIFVCVNAAGDKVGSASDGSVCVLDLRK 514

RESULT 2
TBLR MOUSE STANDARD: PRT; 514 AA.
AC Q8BHJ5; Q8CEG4; Q8VEG3; Q9EQD4;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE F-box-like/WD-repeat protein TBLR1 (Nuclear receptor corepressor/HDAC3
DE complex subunit TBLR1) (TBL1-related protein 1).
GN Name=Tblr1; Synonyms=Iral;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Zhang X., Dormady S., Baech R.;
RT "Identification of four human cDNAs that are differentially expressed
RT by early hematopoietic progenitors.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Testis;
RX MEDLINE=22354683; PubMed=12456851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanpin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmer S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wymshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Iotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sakaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RN Nature 420:563-573 (2002).
RP [3]
RP SEQUENCE OF 314-514 FROM N.A.
RC TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

```
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
CC transcription units. Plays an essential role in transcription
CC activation mediated by nuclear receptors. Probably acts as
CC integral component of the N-CoR corepressor complex that mediates
CC the recruitment of the 19S proteasome complex, leading to the
CC subsequent proteosomal degradation of N-CoR complex, thereby
CC allowing cofactor exchange, and transcription activation (By
CC similarity).
CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least
CC composed of NCOR1, NCOR2, HDAC3, TBLX1, TBLX2, and GPC2.
CC Probable component of some E3 ubiquitin ligase complex. Interacts
CC with histones H2B and H4 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes (By similarity).
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -!- SIMILARITY: Contains 1 F-box-like domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: AF268195; AAG44738.1; -
EMBL: AK029595; BAC26526.1; -
EMBL: AK033347; BAC28241.1; -
EMBL: AK036064; BAC29294.1; -
EMBL: BC018512; AAH18512.1; -
HSSP: P16649; 1ERJ.
DR InterPro: IPR006594; Lish.
DR InterPro: IPR011047; Quin_alc_DH_like.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00120; WD40; 8.
DR PROSITE: PS50896; Lish; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 4.
DR PROSITE: PS00682; WD_REPEATS_2; 6.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR Activator; Nuclear protein; Repeat; Transcription regulation;
KW Ub1 conjugation pathway; WD repeat.
FT DOMAIN 4 36 Lish.
FT DOMAIN 41 86 F-box-like.
FT REPEAT 167 206 WD 1.
FT REPEAT 223 262 WD 2.
FT REPEAT 264 303 WD 3.
```

```
FT REPEAT 306 344 WD 4.
FT REPEAT 347 386 WD 5.
FT REPEAT 389 437 WD 6.
FT REPEAT 440 479 WD 7.
FT REPEAT 481 513 WD 8.
FT DOMAIN 108 117 Poly-Ala.
FT CONFLICT 185 185 A -> V (in Ref. 1).
FT CONFLICT 374 374 X -> R (in Ref. 2; BAC29294).
FT CONFLICT 480 480 Y -> C (in Ref. 3).
SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F9BF14 CRC64;

Query Match 98.9%; Score 3618; DB 1; Length 514;
Best Local Similarity 98.6%; Pred. No. 7.5e-65;
Matches 507; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MSISDEVNVLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNVLVRYLQESGFSHSAFTGIESHSQSNINGALVPPAALISIIQKGLQYV 60

Qy 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQOAAAAAASQ 120
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQOAAAAAATNQ 120

Qy 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFCANVPV 180
Db 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFCANVPV 180

Qy 181 SLLASGSGSTARIWNLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLDWNSGTLT 240
Db 181 SLLASGSGSTARIWNLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLDWNSGTLT 240

Qy 241 ATGSYDGFARITWKDGNLSTLQHGKGPFIKALKNKNGFILSAGVDKTTIIWDAHTGEA 300
Db 241 ATGSYDGFARITWKDGNLSTLQHGKGPFIKALKNKNGFILSAGVDKTTIIWDAHTGEA 300

Qy 301 KQGFPPHSAFALVDVMSNNTFASCTDMCIHVCKLQDQRPDKTFOGHTNEVNAIKWDPT 360
Db 301 KQGFPPHSAFALVDVMSNNTFASCTDMCIHVCKLQDQRPDKTFOGHTNEVNAIKWDPT 360

Qy 361 GNLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGPTNNPNANMLASAF 420
Db 361 GNLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGPTNNPNANMLASAF 420

Qy 421 DSTVRLWDVDRGCIHTLTKHQPVSVAFPSDGRVYLASGFKCVHIWNTQTGALVHSY 480
Db 421 DSTVRLWDVDRGCIHTLTKHQPVSVAFPSDGRVYLASGFKCVHIWNTQTGALVHSY 480

Qy 481 RTGGIPEVCWNAAGDKVGSASDGSVCVLDLRK 514
Db 481 RTGGIPEVCWNAAGDKVGSASDGSVCVLDLRK 514
```

RESULT 3

```
Q6GPC6 ID Q6GPC6 PRELIMINARY; PRT; 522 AA.
AC Q6GPC6;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 03-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX Strausberg R.L., Reingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whaley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Kzyvinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
initative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.L., Pontius J., Clifton S.W.,
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 8 WD repeats.
DR EMBL: BC073215; AAH73215.1; -.
DR InterPro: IPR006594; Lish.
DR InterPro: IPR011047; Quin_alc_DH_like.
DR InterPro: IPR001680; WD40.
DR PRINTS: PR00320; GPROTEINBRPT.
DR Pfam: PF00400; WD40; 8.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00320; WD40; 8.
DR PROSITE: PS00896; LISH; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 4.
DR PROSITE: PS00824; WD_REPEATS_2; 6.
DR PROSITE: PS00294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 522 AA; 56284 MW; 4FB0F58F503B461F CRC64;

Query Match 98.1%; Score 3589.2; DB 2; Length 522;
Best Local Similarity 96.4%; Pred. No. 3.2e-64;
Matches 503; Conservative 5; Mismatches 6; Indels 8; Gaps 3;

Qy 1 MSISDEVNVLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNVLVRYLQESGFSHSAFTGIESHSQSNINGALVPPAALISIIQKGLQYV 60

Qy 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQO-----AAAAAAA 116
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQOAAAAAAA 120

Qy 117 AA-S---QGSARKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEV 172
Db 121 AATSANNQPPAKNGESTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEV 180

Qy 173 FICANNPVSDLLASGSGSTARIWNLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLD 232
Db 181 FICANNPVSDLLASGSGSTARIWNLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLD 240

Qy 233 WNSEGTLLATGSDGFARITWKDGNLSTLQHGKGPFIKALKNKNGFILSAGVDKTTII 292
Db 241 WNSEGTLLATGSDGFARITWKDGNLSTLQHGKGPFIKALKNKNGFILSAGVDKTTII 300

Qy 293 WDAHTGEAKQPPHSAFALVDVMSNNTFASCTDMCIHVCKLQDQRPDKTFOGHTNEV 352
Db 293 WDAHTGEAKQPPHSAFALVDVMSNNTFASCTDMCIHVCKLQDQRPDKTFOGHTNEV 352
```

```
Db 301 WDAHTGEAKQPPHSPALPDVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEV 360
Qy 353 NAIKWDPTGNLLASCSDMTLTKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNAN 412
Db 361 NAIKWDPTGNLLASCSDMTLTKIWSMKHDTICVHDLQAHNKEIYTIKWSPTGPGTNNPNAN 420
Qy 413 LMLASAFDSTVRLWDVDRGICHTLTKHQEPVYSAFSPDGRYLASGSPDKCVHIWNTQ 472
Db 421 LMLASAFDSTVRLWDVDRGICHTLTKHQEPVYSAFSPDGRYLASGSPDKCVHIWNTQ 480
Qy 473 TGAHVSYRGTTGGIFEVCAAGDKVGSASDGSVCVLDLRK 514
Db 481 TGAHVSYRGTTGGIFEVCAAGDKVGSASDGSVCVLDLRK 522

RESULT 4
Q7SZM9 PRELIMINARY; PRT; 519 AA.
ID Q7SZM9 ID Q7SZM9 STANDARD; PRT; 527 AA.
AC Q7SZM9 AC Q7SZM9 Q8BYQ4; Q8COA1;
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA PubMed=12794076;
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.;
RT "Fusion Protein of Retinoic Acid Receptor (alpha) with Promyelocytic
RT Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein
RT Recruits N-CoR-TBLR1 Corepressor Complex to Repress Transcription in
RT Vivo.";
RL J. Biol. Chem. 278:30788-30795 (2003).
CC -1- SIMILARITY: Contains 8 WD repeats.
DR EMBL; AY225098; AAP20646.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR006594; L1SH.
DR Pfam; PF00400; WD40; 8.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; L1SH; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; L1SH; 1.
DR PROSITE; PS00678; WD REPEATS_1; 4.
DR PROSITE; PS00082; WD REPEATS_2; 6.
DR PROSITE; PS0294; WD REPEATS_REGION; 1.
KW Receptor; Repeat; WD-repeat.
SQ SEQUENCE 519 AA; 56043 MW; 5E998EDC8C92296 CRC64;

Query Match 97.9%; Score 3580.5; DB 2; Length 519;
Best Local Similarity 96.5%; Pred. No. 4.8e-64;
Matches 501; Conservative 7; Mismatches 6; Indels 5; Gaps 2;

Qy 1 MSISSEVNLVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISSEVNLVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Qy 61 EAEVSNEDGTLFDGPRIEISLIDAVPDVDTQQAQYRDKLAQQQ--AAAAA 118
Db 61 EAEVSNEDGTLFDGPRIEISLIDAVPDVDTQQAQYRDKLAQQQATAAAAAA 120
Qy 119 S---QQGSAKNGENTANGEANGAHTIANNHTDMMEYDGDVIEIPPKAVLRLGHSEVFIC 175
Db 121 TPNNQPPAKNGENTANGEANGGHALANNHTDMMEYDGDVIEIPSKAVLRLGHSEVFIC 180
Qy 176 ANPVSDDLASGSGDSTARINWLSNSTSGSTQVLRLHCHIREGQDVPSNKDVTSLDWNS 235
Db 181 ANPVSDDLASGSGDSTARINWLSNSTSGSTQVLRLHCHIREGQDVPSNKDVTSLDWNS 240
```

```
Qy 236 EGTLLATGSDYGFARITWTKDGNLSTLQHGKGPFIKWKNNKGNFILSAGVDKTTIIWDA 295
Db 241 EGTLLATGSDYGFARITWTKDGNLSTLQHGKGPFIKWKNNKGNFILSAGVDKTTIIWDA 300
Qy 296 HTGEAKQPPHSPALPDVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAI 355
Db 301 HTGEAKQPPHSPALPDVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAI 360
Qy 356 KWDPTGNLLASCSDMTLTKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTNNPNANLML 415
Db 361 KWDPTGNLLASCSDMTLTKIWSMKHDTICVHDLQAHNKEIYTIKWSPTGPGTNNPNANLML 420
Qy 416 ASAFDSTVRLWDVDRGICHTLTKHQEPVYSAFSPDGRYLASGSPDKCVHIWNTQTA 475
Db 421 ASAFDSTVRLWDVDRGICHTLTKHQEPVYSAFSPDGRYLASGSPDKCVHIWNTQTA 480
Qy 476 LVHSYRGTTGGIFEVCAAGDKVGSASDGSVCVLDLRK 514
Db 481 LVHSYRGTTGGIFEVCAAGDKVGSASDGSVCVLDLRK 519

RESULT 5
TBLX MOUSE
ID TBLX MOUSE STANDARD; PRT; 527 AA.
AC Q9QX87; Q8BMM0; Q8BYQ4; Q8COA1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein TBLX (Transducin beta-like 1X protein).
GN Name=Tblx; Synonyms=Tb11;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J;
RA TISSUE=Hypothalamus; Medulla oblongata; and Pituitary;
RA MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaio I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla B., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Haehizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6; TISSUE=Brain;
RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altechul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Dicchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Stapleton M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fanej J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 200-527 FROM N.A.
 RC STRAIN-129/SvJ;
 RA Bocherby M.R.M., Straw R., Clarke D., Greystrom J.S., Weston P.,
 RA Hunter G., Kimberley C., Rhodes M.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=99264241; PubMed=10330347;
 RA Bassi M.T., Ramasar R.S., Caciotti B., Winship I.M., De Grandi A.,
 RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.;
 RT "X-linked late-onset sensorineural deafness caused by a deletion
 RL involving OAI and a novel gene containing WD-40 repeats.";
 CC Am. J. Hum. Genet. 64:1604-1616(1999).
 CC -!- FUNCTION: F-box-like protein involved in the recruitment of the
 CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
 CC transcription units. Plays an essential role in transcription
 CC activation mediated by nuclear receptors. Probably acts as
 CC integral component of corepressor complexes that mediates the
 CC recruitment of the 19S proteasome complex, leading to the
 CC subsequent proteosomal degradation of transcription repressor
 CC complexes, thereby allowing cofactor exchange (By similarity).
 CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least
 CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.
 CC Component of an E3 ubiquitin ligase complex containing UBE2D1,
 CC STAH1, CACYPB/STP, SKP1A, APC and TBL1X. Probably part of other
 CC corepressor complexes, that do not contain NCOR1 and NCOR2.
 CC Interacts with histones H2B, H3a and H4 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- TISSUE SPECIFICITY: Expressed in the cochlea.
 CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and
 CC apparently displays the same function as component of ubiquitin E3
 CC ligase complexes (By similarity).
 CC -!- SIMILARITY: Belongs to the WD-repeat E3 family.
 CC -!- SIMILARITY: Contains 1 F-box-like domain.
 CC -!- SIMILARITY: Contains 8 WD repeats.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; AK030547; BAC27015.1; -;
 CC DR EMBL; AK031937; BAC27612.1; -;
 CC DR EMBL; AK038674; BAC30092.1; -;
 CC DR EMBL; BC043105; AAH43105.1; -;
 CC DR EMBL; F38006; CAB61534.1; ALT_SEQ.
 CC MGD; MGI:1336172; Tbl1x.

DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 5.
 DR PRINTS: PR00320; GPROTEINRPT.
 DR ProDom: PD000018; WD40; 2.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS00896; LISH; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; 4.
 DR PROSITE: PS00082; WD_REPEATS_2; 6.
 DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
 KW Activator; Nuclear protein; Repeat; Transcription regulation;
 KW Ub1 conjugation pathway; WD repeat.
 FT DOMAIN 4 36 Lish.
 FT DOMAIN 41 86 F-box-like.
 FT REPEAT 180 219 WD 1.
 FT REPEAT 236 275 WD 2.
 FT REPEAT 277 316 WD 3.
 FT REPEAT 319 359 WD 4.
 FT REPEAT 360 399 WD 5.
 FT REPEAT 402 450 WD 6.
 FT REPEAT 453 492 WD 7.
 FT REPEAT 494 526 WD 8.
 FT DOMAIN 108 121 Poly-Ala.
 FT DOMAIN 130 133 Poly-Ala.
 FT CONFLICT 104 104 A -> T (in Ref. 1; BAC27015).
 FT CONFLICT 349 349 L -> H (in Ref. 1; BAC27612).
 SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;
 Query Match 93.4%; Score 3418.1; DB 1; Length 527;
 Best Local Similarity 88.1%; Pred. No. 1.4e-60;
 Matches 467; Conservative 29; Mismatches 15; Indels 19; Gaps 8;
 QY 1 MSISDEVNPLVYRYLQESGFSHSAFTGKISHISOSNINGALVPPAALISILQKGLQYV 60
 DB 1 MSISDEVNPLVYRYLQESGFSHSAFTGKISHISOSNINGALVPPAALISILQKGLQYV 60
 QY 61 EAEVSNEDGTLFDGRIEISLSLIDAVMPDVVQTRQAYRDKLAQQA-AAAAAAAAA- 118
 DB 61 EAEISINEDGTVDGRIEISLSLIDAVMPDVVQTRQAYRDKLAQQA-AAAAAAAAA- 120
 QY 119 --S-----QQSAXNGENTANGENGAAHTANNNTDM-MEVDGVEIPPNKAVL 165
 DB 121 ATSTAATTPAAAQQNPKNXGATVNGEENGAAHAI--NNHS-KPMEIDGVEIPPSKATVL 178
 QY 166 RGHSEVFTICANNPVSDDLASGGSTARIWNLSENSTSGTQLVLRHCIRREGGQDVPSN 225
 DB 179 RGHSEVFTICANNPVSDDLASGGSTARIWNLSENSTSGTQLVLRHCIRREGGQDVPSN 238
 QY 226 KDVTSLDWNSEGLLATGSDYGFARITWKDGNLSTLQGHKGPFPALKWNKGNFTLSAG 285
 DB 239 KDVTSLDWNSEGLLATGSDYGFARITWKDGNLSTLQGHKGPFPALKWNKGNFTLSAG 298
 QY 286 VDKTTIINDAHTGEAKQOPFFHSAPALVDVQSN--TFASCSTDMCIHVCKLQDRPIKT 344
 DB 299 VDKTTIINDAHTGEAKQOPFFHSAPALVDVQ--NNTTFASCSTDMCIHVCKLQDRPVKT 357
 QY 345 FQGHTEVNAIKWDPTGNLLASCDMDTLKISWMDKQNCVDLQOHKKEIYTIKWSPTGP 404
 DB 358 FQGHTEVNAIKWDPTGNLLASCDMDTLKISWMDKQNCVDLQOHKKEIYTIKWSPTGP 417
 QY 405 GTNNPNANIMLASGFDSTVRLWDVRGICHTLTKHQEPVYSVAFSPDGRYLASGFDK 464
 DB 418 ATSNPNINMLASGFDSTVRLWDVRGICHTLTKHQEPVYSVAFSPDGRYLASGFDK 477
 QY 465 CVHIWNTQTGALVSHVSGTGGIFEVCMNAGDKVASASDGSVCVLDLRK 514
 DB 478 CVHIWNTGSLVSHVSGTGGIFEVCMNARGDKVASASDGSVCVLDLRK 527
 RESULT 6
 TBLX HUMAN
 ID TBLX HUMAN STANDARD; PRT; 526 AA.
 AC O60907; Q86UY2;
 DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein TBL1X (Transducin beta-like 1X protein)
DE (Transducin-beta-like 1, X-linked) (SWAP55).
CN Name=TBL1X; Synonyms=TBL1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DISEASE.
RX MEDLINE=99264241; PubMed=10330347;
RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,
RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.,
RT "X-linked late-onset sensorineural deafness caused by a deletion
RT involving OAI and a novel gene containing WD-40 repeats.";
RL Am. J. Hum. Genet. 64:1604-1616(1999).
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=Lymph, and Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Diatchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RN IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR
RP COMPLEX WITH NCOR2 AND HDAC3.
RX PubMed=10809664;
RA Guencher M.G., Lane W.S., Fischle W., Verdin E., Lazar M.A.,
RA Shiekhattar R.;
RT "A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-
RT repeat protein linked to deafness.";
RL Genes Dev. 14:1048-1057(2000).
RN [4]
RP COMPONENT OF THE N-COR COMPLEX WITH NCOR2 AND HDAC3.
RX PubMed=10944117;
RA Li J., Wang J., Wang J., Nawaz Z., Liu J.M., Qin J., Wong J.;
RT "Both corepressor proteins SMRT and N-CoR exist in large protein
RT complexes containing HDAC3.";
RL EMBO J. 19:4342-4350(2000).
RN [5]
RP SUBUNIT OF A COMPLEX WITH UBE2D1, CACVBP, SIAH1 AND APC.
RX PubMed=11389839;
RA Matsuzawa S.-I., Reed J.C.;
RT "Siah-1, SIP, and Ebi collaborate in a novel pathway for beta-catenin
RT degradation linked to p53 responses.";
RL Mol. Cell 7:915-926(2001).
RN [6]
RP COMPONENT OF THE N-COR COMPLEX WITH NCOR1, NCOR2, GPS2, TBL1R AND
RP HDAC3.
RX MEDLINE=21930350; PubMed=11931768;
RA Zhang J., Kalkum M., Chait B.T., Roeder R.G.;
RT "The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK
RT pathway through the integral subunit GPS2.";
RL Mol. Cell 9:611-623(2002).
RN [7]

RP COMPONENT OF THE N-COR COMPLEX WITH TBL1R; CORO2A AND HDAC3, AND
RP HISTONE-BINDING.
RX PubMed=12628926;
RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,
RA Wong J.;
RT "Purification and functional characterization of the human N-CoR
RT complex: the roles of HDAC3, TBL1 and TBL1R.";
RL EMBO J. 22:1336-1346(2003).
RN [8]
RN FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.
RX PubMed=14980219;
RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;
RT "A corepressor/coactivator exchange complex required for
RT transcriptional activation by nuclear receptors and other regulated
RT transcription factors.";
RL Cell 116:511-526(2004).
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
CC transcription units. Plays an essential role in transcription
CC activation mediated by nuclear receptors. Probably acts as
CC integral component of corepressor complexes that mediates the
CC recruitment of the 19S proteasome complex, leading to the
CC subsequent proteasomal degradation of transcription repressor
CC complexes, thereby allowing cofactor exchange.
CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least
CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.
CC SIAH1, CACVBP/SIP, SKP1A, APC and TBL1X. Probably part of other
CC corepressor complexes, that do not contain NCOR1 and NCOR2.
CC -!- INTERACTS WITH HISTONES H2B, H3A AND H4.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes (By similarity).
CC -!- DISEASE: Defects in TBL1X may be a cause of late-onset
CC sensorineural deafness (OASD) [MIM:300650]. OASD is an X-linked
CC recessive disorder characterized by ocular albinism and
CC progressive sensorineural hearing loss in the fourth and fifth
CC decades of life. OASD may be caused by deletion of both GPR143/OAI
CC and TBL1X adjacent genes; TBL1X defects possibly causing the
CC hearing phenotype.
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -!- SIMILARITY: Contains 1 F-box-like domain.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y12781; CAA73319.1; ALT_INIT.
DR EMBL: BC032708; AAH32708.1; -
DR EMBL: BC052304; AAH52304.1; ALT_INIT.
DR HSSP: P16649; IERJ.
DR Genew; HGNC:11585; TBL1X.
DR MIM; 300196; -
DR MIM; 300650; -
DR GO; GO:0007605; P:perception of sound; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR GO; GO:0007601; P:visual perception; TAS.
DR InterPro; IPR0011047; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.

```
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00882; WD REPEATS 2; 6.
DR PROSITE; PS00294; WD REPEATS REGION; 1.
KW Activator; Deafness; Nuclear protein; Repeat;
FT DOMAIN 4 36
FT DOMAIN 41 86
FT REPEAT 179 218
FT REPEAT 235 274
FT REPEAT 276 315
FT REPEAT 318 358
FT REPEAT 359 398
FT REPEAT 401 449
FT REPEAT 452 491
FT REPEAT 493 525
FT REPEAT 526 564
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6B9 CRC64;
Query Match 91.7%; Score 3355.2; DB 1; Length 526;
Best Local Similarity 86.2%; Pred. No. 3.1e-59;
Matches 456; Conservative 38; Mismatches 17; Indels 18; Gaps 10;
Qy 1 MSISDSEVFLVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSITSDEVFLVRYLQESGFSHSAFTGIESHISQSNINGTLVPPAALISILQKGLQYV 60
Qy 61 EAEVINEDETLPCRIEISLIDAVMPDVVQTRQAQYRDKLAAQQ--AAAAAAA-AA 117
Db 61 EAEISINEDGTVFDGRIEISLIDAVMPDVVQTRQAQYRDKLAAQQASAAAAATAA 120
Qy 118 ASQ-----QG-S-----AKNGENTANGENGATHTANNHTDM-MEVGDGVIPNKAVALR 166
Db 121 ATAATTSAGVSHQNPKNREAVNGENRAHSV--NNHA-KPMEDGVEIIPSKATVLR 178
Qy 167 GHESEVFIKAWNPVSDLLASGSDSTARIWNLSENSTSGTQLVLRHCIREGGQDVPSNK 226
Db 179 GHESEVFIKAWNPVSDLLASGSDSTARIWNLSENSTSGTQLVLRHCIREGGHVDPSNK 238
Qy 227 DVTSLDWNSGTLATSGYGFARITWKDGNLSTLQHGKPIPALKNWKNKGFILSAGV 286
Db 239 DVTSLDWNTGTLATSGYGFARITWKDGNLSTLQHGKPIPALKNWKNKGNVILSAGV 298
Qy 287 DKTTIWDATGTEAKQFPFHSPALPDVQSN--TFASCTSDMCIHVKLGODRPTKF 345
Db 299 DKTTIWDATGTEAKQFPFHSPALPDVQ--NNTTFASCTSDMCIHVKLGODRPTKF 357
Qy 346 QGHTNEVNAIKWPTGNLLASCDSDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGP 405
Db 358 QGHTNEVNAIKWPTGNLLASCDSDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 417
Qy 406 TNNPNANMLASAFDSFTVLVDVDRGICHTLTKHQPYSVAFSPDGRYLASGSPDKC 465
Db 418 TSNPNANMLASAFDSFTVLVDIERGVCTHTLTKHQPYSVAFSPDGRYLASGSPDKC 477
Qy 466 VHIWNTOTGALVHSYRTGGIFEVCMNAAGDKVCASASDGSVCVLDLRK 514
Db 478 VHIWNTUSGNLVHSYRTGGIFEVCMNARGDKVCASASDGSVCVLDLRK 526
RESULT 7
ID TBL1 HUMAN
AC Q9BQ87; STANDARD; PRT; 522 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DR F-box-like/WD-repeat protein TBL1Y (Transducin beta-like 1Y protein)
DE (Transducin-beta-like 1, Y-linked).
GN Name=TBL1Y; Synonyms=TBL1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND TISSUE SPECIFICITY
MEDLINE=22699787; PubMed=12815422; DOI=10.1038/nature01722;
RA Skaltsky H., Kuroda-Kawaguchi T., Minx P.J., Cordum H.S.,
RA Hillier L.W., Brown L.G., Repping S., Pyntikova T., Ali J., Bieri T.,
RA Chinwalla A., Delehaanty A., Delehaanty K., Du H., Fewell G.,
RA Fulton L., Fulton R., Graves T.A., Hou S.-F., Latrielle P.,
RA Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W.,
RA Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K.,
RA Schultz B., Strong C., Tin-Wollam A., Yang S.-P., Waterston R.H.,
RA Wilson R.K., Rozen S., Page D.C.;
RT "The male-specific region of the human Y chromosome is a mosaic of
discrete sequence classes.";
RL Nature 423:825-837(2003).
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the
ubiquitin/19S proteasome complex to nuclear receptor-regulated
transcription units. Plays an essential role in transcription
activation mediated by nuclear receptors. Probably acts as
integral component of corepressor complexes that mediates the
recruitment of the 19S proteasome complex, leading to the
subsequent proteasomal degradation of transcription repressor
complexes, thereby allowing cofactor exchange (By similarity).
CC -!- SUBUNIT: Probable component of the N-CoR repressor complex and
some E3 ubiquitin ligase complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: Fetal brain and prostate.
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and
apparently displays the same function as component of ubiquitin E3
ligase complexes (By similarity).
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -!- SIMILARITY: Contains 1 F-box-like domain.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF332220; AAK13472.1; -.
EMBL; AF332221; AAK13473.1; -.
EMBL; AF332222; AAK13474.1; -.
HSP; P16649; IERJ.
MIM; 400033; -.
Genew; HGNC:18502; TBL1Y.
InterPro; IPR006594; Lish.
InterPro; IPR001680; WD40.
InterPro; IPR011046; WD40_like.
Pfam; PF00400; WD40; 8.
PRINTS; PR00320; GPROTEINRPT.
ProDom; PD000018; WD40; 3.
ProDom; SM00667; Lish; 1.
SMART; SM00320; WD40; 8.
PROSITE; PS00896; LISH; 1.
PROSITE; PS00678; WD REPEATS 1; 4.
PROSITE; PS00882; WD REPEATS 2; 6.
PROSITE; PS00294; WD REPEATS REGION; 1.
KW Activator; Nuclear protein; Repeat; Transcription regulation;
FT DOMAIN 4 36
FT DOMAIN 41 86
FT REPEAT 179 216
FT REPEAT 233 272
FT REPEAT 274 313
FT REPEAT 316 354
FT REPEAT 357 396
FT REPEAT 399 447
FT REPEAT 450 489
FT REPEAT 491 521
```


RA MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis J.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fletschmann W.,
RA Finkler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195 (2000).
RA [3]
RA SEQUENCE FROM N.A.
RA STRAIN=Berkeley; TISSUE=Embryo;
RA MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Konnoller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celiker S.E.,
RA "A Drosophila full-length cDNA resource."
RA Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
RA [4]
RA FUNCTION, INTERACTIONS WITH PHYL AND SINA, AND MUTANTS EBIC1 AND
RA EBIC3.
RA PubMed=11032805;
RA Boulton S.J., Brook A., Staehling-Hampton K., Heitzler P., Dyson N.,
RA "A role for Ebi in neuronal cell cycle control."
RA EMBO J. 19:5376-5386 (2000).
RA [5]
RA FUNCTION, AND INTERACTIONS WITH SNO; SMR AND SU(H).
RA PubMed=12230979;
RA Tsuda L., Nagaraj R., Zipursky S.L., Banerjee U.,
RA "An EGR/Ebi/Sno pathway promotes delta expression by inactivating
RA Su(H)/SMR repression during inductive notch signaling."
RA Cell 110:625-637 (2002).
RA CC -!- FUNCTION: F-box-like component of E3 ubiquitin ligase complexes;
RA involved in R7 photoreceptor cell differentiation, cone cell
RA development and neuronal cell cycle control. E3 ubiquitin ligase
RA complexes mediate ubiquitination and subsequent proteasomal
RA degradation of target proteins. Required for specification of R7
RA photoreceptor cell fate in the eye by participating to the
RA ubiquitination and subsequent proteasomal degradation of Tramtrack
RA (ttk), a general inhibitor of photoreceptor differentiation.
RA Required to block the S phase entry in the peripheral nervous
RA system and central nervous system in a process that does not
RA involve the degradation of ttk. Involved in cone cell development

CC by preventing the transcriptional repression mediated by Su(H) on
CC DL, probably by participating to a E3 complex that contains sno
CC and mediates the ubiquitination and subsequent proteasomal
CC degradation of some component of the Su(H) repressor complex.
CC -!- SUBUNIT: Component of some E3 complex at least composed of sina,
CC ebi and phyl, required for the degradation of ttk. Probably forms
CC a E3 complex with sno, required for the degradation of some
CC component of the Su(H) repressor complex. Interacts with sno and
CC Su(H) and Smr.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Widely expressed both in embryos and larvae.
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes.
CC -!- MISCELLANEOUS: 'Ebi' means 'shrimp' in Japanese.
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -!- SIMILARITY: Contains 1 F-box-like domain.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
CC frameshift in position 484.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AFI46345; AAD35017.1; --
CC ENBL; A8003589; AAF51501.1; --
CC ENBL; A1061326; AAL28874.1; ALT_FRAME.
CC HSSP; P16649; 1ERJ.
CC FlyBase; FBgn0023444; ebi.
CC GO; GO:000074; P:regulation of cell cycle; IMP.
CC InterPro; IPR006594; Lish.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 3.
CC PRINTS; PR00320; GPROTEINRPT.
CC PRODOM; PD000018; WD40; 3.
CC SMART; SM00667; Lish; 1.
CC SMART; SM00320; WD40; 8.
CC PROSITE; PS00896; Lish; 1.
CC PROSITE; PS00678; WD_REPEATS_1; 3.
CC PROSITE; PS00882; WD_REPEATS_2; 6.
CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
CC Cell cycle; Nuclear protein; Repeat; Ub1 conjugation pathway; Vision;
KW WD repeat.
FT DOMAIN 4 36 Lish.
FT DOMAIN 41 84 F-box-like.
FT REPEAT 353 392 WD 1.
FT REPEAT 408 447 WD 2.
FT REPEAT 449 488 WD 3.
FT REPEAT 491 531 WD 4.
FT REPEAT 532 571 WD 5.
FT REPEAT 574 622 WD 6.
FT REPEAT 625 664 WD 7.
FT REPEAT 666 700 WD 8.
FT DOMAIN 206 323 Ser-rich.
FT MUTAGEN 16 16 L->Q: In ebiCC1; induces ectopic S phases
FT within the peripheral and central nervous
FT system.
FT C->Y: In ebi4; induces defects in R7 cell
FT development in the eye.
FT S->L: In ebiCC3; induces ectopic S phases
FT within the peripheral and central nervous
FT system.
SQ SEQUENCE 700 AA; 72387 MW; 28C6D8D07BB79FB7 CRC64;
Query Match 79.7%; Score 2917.1; DB 1; Length 700;
Best Local Similarity 57.4%; Pred. No. 1.2e-49;
Matches 411; Conservative 50; Mismatches 36; Indels 219; Gaps 58;


```
QY 1 MSISDEVNPLVYRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
D 1 MSPSSDEVNPLVYRYLQESGFLHSAVVGFIHSHISQSNINGALVPPAALLTIQKGLLT 60
QY 61 EAEVSIINEDGTLFD-CRPTESLSLIDAVMPDV-----VQTRQO-----AYRD---KLA-- 104
D 61 EVESVSGEDG---EVARPIEGLSLIDAVMPDVKPLPIVKT-EPGKPGAV-DSSAP-AGG 114
QY 105 -Q-----Q-----Q-A-----AA-A-----A 112
D 115 NQNNNAKPEIKIEPGTGVAGSAGCNKIAGSTGTGTSTPTDOSASEVDSSGNAANAGGTYA 174
QY 113 ----A-----A-----AA-ASQ-----Q-----GSAK--N-G-EN---T-- 130
D 175 GNGAGGNGQASTGGSNSTSTPPAGGDLAAPGASQKKSQNSNEAGSSSGNAGNANATSTD 234
QY 131 -A-----NG-----EE-----N-----G-AH-----TI- 141
D 235 AASSTNGNSTSSSVEQPTSLTTPAGGTVTSNPDAAAGGASTATGSKAPSGAVTIR 294
QY 142 --A--NN-----H-----T-DWMEV--DGD--VEIPNKAIVLRG 167
D 295 VQAQNNVQSGSSNAOSSAPSGTISSTSGGACTPAAL--VPMDIDENIIPESKARVLRG 353
QY 168 HESEVFICAWNPNVS-DLLASGGSDSTARIWNLSE-NSTSGSTQVLRLHCIREGGQDVPSN 225
D 354 HESEVFICAWNPN-SRDLASGGSDSTARIWMSDAN-T-NSNQLVLRHLCIQKGAEPVSN 410
QY 226 KDVTSLDWSNSEGILLATGSDYGPARIW-TYDGNLSTLQGHKGPFPALKKNNK-GNFIUS 283
D 411 KDVTSLDWNCDGSLLATGSDYGVARIWKT-DGRLASTLQGHKGPFPALKKNN-KCGNYILS 468
QY 284 AGVDKTTIIWDATGE-AKQRPFPHPALPDVDQSNNTFASGSD-MCHVCKLG--QD 339
D 469 AGVDKTTIIWDASTGQCT-QQAFHSPALPDVDQNTQAFASGSDTQD-IHVCRLGVNE- 525
QY 340 RPIKTPQGHTEVNAIKWDPGTGNLLASCDDMTLKIWSMKQD-NCVHDLQHNKEIYTIK 398
D 526 -PIKTPKGTNEVNAIKWCPQGLLASCDDMTLKIWSMNRDCC-HDLQAHSEIYTIK 583
QY 399 WSPGTGCTNNPNANMLASGDSFSTVRLWDVDRGICIHITLTKHQEPVYVAFSPDGRYLA 458
D 584 WSPGTGCTNNPNANMLASGDSFSTVRLWDVDRGICIHITLTKHQEPVYVAFSPDGKHLA 643
QY 459 SSGFDCVCHVNTQTGALVHSGYGTGIEFCVNCNNAAGDKVGASASDGSCVCLDLRK 514
D 644 SSGFDCVCHVNTQTGALVHSGYGTGIEFCVNCNNAAGDKVGASASDGSCVCLDLRK 699

RESULT 10
Q96RW7
ID Q96RW7 PRELIMINARY; PRT; 5636 AA.
AC Q96RW7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hemicientin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Trent J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156100; AAK68690.1; -.
DR HSSP; P07996; 1LSL.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005021; F:vascular endothelial growth factor receptor. .; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
```

```
DR InterPro; IPR000875; Cecropin.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006605; G2F.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR000169; Pept_cys_acsite.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR009134; VEGFR.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF07645; EGF_CA; 8.
DR Pfam; PF07474; G2F; 1.
DR Pfam; PF00047; Ig; 44.
DR Pfam; PF00090; TSP_1; 6.
DR PRINTS; PR01832; VEGFRECEPTOR.
DR SMART; SM00179; EGF_CA; 7.
DR SMART; SM00408; IGC2; 43.
DR SMART; SM00209; TSP1; 6.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00268; CECROPIN; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS00266; EGF_3; 5.
DR PROSITE; PS01187; EGF_CA; 8.
DR PROSITE; PS00835; IG_LIKE; 44.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
DR PROSITE; PS00092; TSP1; 6.
DR EGF-like domain.
KW
SQ SEQUENCE 5636 AA; 613673 MW; F00B319CED7B52C CRC64;
```

Query Match 71.5%; Score 2614.8; DB 2; Length 5636;

Best Local Similarity 9.5%; Pred.No. 1.4e-41;

Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S-----I-----SS---D-----EV-----8

D 1 MISWEVHTVPLFALLYSSLAQDASPOSIRAEPEPEGASTLAFVDFVTGSMYDDLQVVI 60

QY 9 -----NP-LV-----12

D 61 EGASKILETSLKRPKPLNFALVFPHPDPEIGFVTTTDPKRFQYBELRELYVQGGDCPE 120

QY 13 -----Y---R---Y---LQ-----17

D 121 MSIGAIAKTALEISLPGSFYVFTDARSKDYRLTHEVLQLIQKQSQVFLTGDCDDRTH 180

QY 18 -----E---SG--F-----S--H--S-----A---F 26

D 181 IGYKVVEEIASTSSQGVFHLDDKKQVNEVLKQVVEAAQASKVHLLSTDHLEQAVNTWRIFP 240

QY 27 -----T-----FG-----I-----K 31

D 241 DPSLKEVTLSGSPSMIEIRNPLGKLKKGFGLHELLIHNSAKVNVNKEPEGAMWTVK 300

QY 32 ---S--H-----I-----S---Q-----36

D 301 TSSSGRHVRITGLSTIDFRAGFSRKPTLDFKFTVSRPVQGIPTVYVLLNTSGISTPARID 360

QY 37 -----S-----N-----G-----41

D 361 LLELLISGSSLUKTIPIVKYYPHKKPYGIWNISDFVPPNEAFPLKVTGYDKDDYLFQVSS 420

QY 42 -----A-----L---VP-----P-----A 47

D 421 VSFSSIVPDAPKVTMPKTPGYLQPGQIPCSVDSLPPFTLSFVRNGVTLGVDQYLKESA 480

QY 48 -----A---L-----I-----S---IIO-----KG-----56

D 481 SVSLDIAKVTLSDEGFYECIAVSSAGTGRAQTFFDVSEPPPIQVNNVTVTPGERAVLT 540

QY -57 -----L--Q-----Y---V--E-----61

541 DB CLIIASVDYNLWQRNDRVRLAEPARIRTLANLSLEKSVKFNDAEGYHCVMSSEGGSS 600
542 QY -A- - - - -EVS- - - - -NE- - - - - 68
601 DB AASVFLTVQBPVKVVMKQNSFTGGSEVMSGATGPKPKIAWTVNDMFIIVGSHRYRM 660
602 QY -D- - - - -G- - - - -R- - - - -P- - - - -I- - - - - 78
661 DB TSDGTLFIKNAAPKADAGIYGLASNSAGTDKQNSTLYIEAPKLMVVOSELLVALGDTIV 720
79 QY -E- - - - -L- - - - -S- - - - -LI- - - - -DA- - - - -V- - - - - 87
721 DB MECKTSGIPPPQKWKFGDLELSPSTELIDPLGLLKIQTQDLDAGDYTCVAINAAGR 780
88 QY -M- - - - -P- - - - -DV- - - - -VQ- - - - -T- - - - - 94
781 DB ATGKITLDVGSPPVFIQEPADVSMIEGNSVTLFCYVQGYPEPTIKWRRLDNMFIISRPF 840
95 QY -R- - - - -R- - - - -R- - - - -R- - - - -R- - - - - 95
841 DB VSSISQRLTGALFILNLWASDKGYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVAN 900
96 QY -Q- - - - -Q- - - - -Q- - - - -Y- - - - -A- - - - -R- - - - -D- - - - -KL- - - - - 103
901 DB VIEGQQLTLPCTLLAGNPIPERRWIKNSAMLLQNPYITVRSGLSHIERVOLQDGGEYTC 960
104 QY -A- - - - -A- - - - -Q- - - - -Q- - - - -AA- - - - - 109
961 DB VASVAGTNNKTTVVVHVHVLPTIHQGOILSTIEGIVTLPCKASGNPKPSVWSKKGEL 1020
110 QY -AAA- - - - -A- - - - -A- - - - -A- - - - -A- - - - - 118
1021 DB ISTSSAKFSAGAGSLYVSPGESEGYVCTAINTAGYAKRKVQLTVYVPRVFGDLRG 1080
119 QY -SQ- - - - -S- - - - -Q- - - - -Q- - - - -GS- - - - - 123
1081 DB LSQDKPVEISVLAGEEVLTPCEVKSLLPPIITWAKETQLISPFSPRHTLPSPGSMKITET 1140
124 QY -A- - - - -A- - - - -K- - - - -N- - - - -N- - - - - 126
1141 DB RTSDSGMYLCVATNIAGNVQAVKLVNHVPPKIQRGPKHLKVQGVQVDPICNAQGTPLP 1200
127 QY -GE- - - - -N- - - - -T- - - - -A- - - - -N- - - - -G- - - - - 133
1201 DB VITWSKGSMTLVDGHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260
134 QY -E- - - - -E- - - - -N- - - - -GA- - - - -H- - - - - 139
1261 DB PPTVEDLEPPYNTTFQERVANQRIEFPKPAKGTPTIKWLHNGRELTGREPGISILEEG 1320
140 QY -T- - - - -IA- - - - -N- - - - -N- - - - -HT- - - - -DM- - - - - 148
1321 DB TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKDKBQVSNVSVLLNQLTNL 1380
149 QY -M- - - - -M- - - - -E- - - - -E- - - - - 150
1381 DB FCEVEGTPSPILIMWKONVQVTESSITQTVNNGKILKLFRATPEDAGRYSCKAINIAGTS 1440
151 QY -VD- - - - -G- - - - -DV- - - - -E- - - - -I- - - - -P- - - - -PN- - - - - 160
1441 DB QKYNIDVLVPPTIIGTNFPKEVSVNLNRDVALEQVKGTPFPDIHWFKDGKPLFLGDPN 1500
161 QY - - - - - - - - - -K- - - - -A- - - - -V- - - - -VL- - - - - 165
1501 DB VELLDRGQVLHLKNARNNDKGRYOCVTVSNAAGQAKDKITITVPPSIKGNVTDTISVL 1560
166 QY -RG- - - - -RG- - - - -H- - - - -H- - - - -H- - - - - 168
1561 DB INSLIKLECKTRGLPMPAITWYKDGQPIMSSQALYIDKQYLHIPRAQVSDSATYTCV 1620
169 QY -E- - - - -S- - - - -E- - - - -VFI- - - - -C- - - - -A- - - - -W- - - - - 177

1621 DB ANVAGTAESFHVVDVVPVPMIEGNLATPLNKQVVIASHSLTECNAAGNPSILTLWKDGV 1680
178 QY -N- - - - -N- - - - -P- - - - -V- - - - -SD- - - - - 182
1681 DB PVKANDNFRIAGGKKLEIMSAQIEDRGQVIVCATSVAGEKEIKYEVDVLPVPAIEGGDE 1740
183 QY -LL- - - - -ASGS- - - - -G- - - - -D- - - - -S- - - - - 191
1741 DB TSYFIVVNNLLELDCHVTGSPPTIMLKDQGLIDDERDGFKILLNKRKLIVIAQAQVSNT 1800
192 QY -TA- - - - -R- - - - -R- - - - -I- - - - -W- - - - - 196
1801 DB GLYRCMAANTAGDHKEFEVTHVPTTIKSSGLSERVWVKYKPVALQCIANGIPNSITW 1860
197 QY -NL- - - - -NL- - - - -S- - - - -S- - - - -S- - - - - 199
1861 DB LKDDQPVNTAQGNLKIQSSGRVLQIAKTLLLEDAGRYTCVATNAAGETOQHILHVHEPPS 1920
200 QY -E- - - - -N- - - - -S- - - - -T- - - - -SGST- - - - - 207
1921 DB LEDAGRMLETVLVSNPVLQECKAAGNPVPTIWTYKDNCLLSGSTMTFLNRQIIDIES 1980
208 QY -Q- - - - -L- - - - -V- - - - -LR- - - - -H- - - - -C- - - - -I- - - - - 215
1981 DB AQISDAGIYKCVAINASAGATELFYSLQVHVAFPSISGSNNMVAVVNNVPRLECEARGIPA 2040
216 QY - - - - -R- - - - -E- - - - -G- - - - -G- - - - -Q- - - - -D- - - - - 221
2041 DB PSUTWLKOGSPVSSFSNGLOVLGGRIALATSTQISDTGRYTCVAVNAAGEKORDILRV 2100
222 QY -VPSN- - - - -KD- - - - -VT- - - - -S- - - - -L- - - - - 231
2101 DB YPPNIMBEEQNVSVLISOAVELLQSDAI PPPTTLTWLKDGHPLKKPGLSISENSVLK 2160
232 QY -D- - - - -D- - - - -W- - - - -N- - - - -S- - - - -EG- - - - -TLL- - - - - 240
2161 DB IEDAQVQDTGRYTCEATNVAGTEKKNVNNVWPPNIGSGDELQTLTVEGNLISLCE 2220
241 QY -AT- - - - -GS- - - - -Y- - - - -D- - - - -G- - - - -F- - - - -A- - - - - 249
2221 DB SSGIPPNLIWKKGSFVLDSMGRVRLSGRQLOISAEKSDAALYSCVASNAGTAK 2280
250 QY - - - - -R- - - - -I- - - - -I- - - - -WTKDG- - - - - 256
2281 DB KEYNLQVIRPTITNSGSHPTLIIVTRGKISISLECEVOGIPPTVTWMDGHPHAKGV 2340
257 QY -NL- - - - -AS- - - - -T- - - - -L- - - - -L- - - - -GOH- - - - - 265
2341 DB EILDEGHILQKNHIVSDTGRYCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400
266 QY -K- - - - -G- - - - -PI- - - - -FALK- - - - -W- - - - - 274
2401 DB EKNVSLTCEASGIPLPSTWTF--KDGWPVLSNSVRILSGGRMLRLMQTTMEDAQYTC 2458
275 QY - - - - -N- - - - -K- - - - -K- - - - -GN- - - - - 279
2459 DB VVRNAGEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPVITWHKDG 2518
280 QY -F- - - - -I- - - - -I- - - - -L- - - - -S- - - - -AGV- - - - - 286
2519 DB QPLODEAHIIISGGRFLQITNVQVPHGTGRYTCLASSPAGHKRSFSLNVFVSPTIAGVG 2578
287 QY -D- - - - -K- - - - -T- - - - -I- - - - -I- - - - -W- - - - - 293
2579 DB SGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLESNRNRIILPGRTQLILNAQ 2638
294 QY -D- - - - -A- - - - -H- - - - -H- - - - -TG- - - - -E- - - - - 299
2639 DB EDNAGRYSCVATNEAGEMIKHYEVKVYIPIINKGDLWGLSPKEVKIKVNNLTLECE 2698
300 QY -A- - - - -K- - - - -K- - - - -Q- - - - -Q- - - - - 303
2699 DB AYAIPSASLSWYKDGOPKSDDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELD 2758

Db 4381 AAPVIRKIBPLEVALGHLAKFTCEIQSAPNRFQWFKAGREIYESDKCSIRSKYISSL 4440
QY 134 E-----E-----N----- 136
Db 4441 EILRTQVDCGEYTCASNEYGSVCATLVTVEAYPTPLSRPKSLTTFVGKAAFICT 4500
QY 137 -----GA-----H-----TI-----A-N----- 143
Db 4501 VTGTPVETIWDGAALSPNWKLSAENKHILELSNLTIDRGVYSCASNKFCADI 4560
QY 144 -----N-----H----- 145
Db 4561 COAELIIDKPHFIKELEPVQSAINKVHLEQVDEDRKVTVTWSDGQKLPFGDYKIC 4620
QY 146 -----T----- 148
Db 4621 FEDKIATLEIPLAKLDKSGTYVCTASNEAGSSCSATVTVREPPSFVKYKVDPSYMLPGE 4680
QY 149 -----M-----E-----V-D-G----- 153
Db 4681 SARLHCKLKGSPVIQVTFKNNKELSENTVMYFVNSEAILDITDVKVEDSDSYCEAV 4740
QY 154 -DV-----EI-----PP----- 159
Db 4741 NDVGSDSCSTEIVIKEPPSPFKTLEPADIVRGTNALLQCEVSGTGPFEISWFKDKQIRS 4800
QY 160 -----N-----K----- 161
Db 4801 SKYRIFSOKSLVLEIFSPNSADVGECVANEVKGCGCMATHLLKBPPTFVKVDDL 4860
QY 162 -----AV-----VL----- 165
Db 4861 IALGGQVTLQAAVRGSEPISTVMKQEVIREDEGKIKMSFNGVAVLIIPDVQISFGGK 4920
QY 166 -----R----- 166
Db 4921 YTCLAENEAGSQTSGELIVKEPAKIERAELIQVTAGDPATLEYTVAGTPELKPKWKYD 4980
QY 167 G-----H-----E-S-EV-----F-----I-- 174
Db 4981 GRPLVASKKYRISFKNNVAQLKPYSAELHDGQYTFEISNEVGSSCETTFVLDRDIAP 5040
QY 175 -----C-A-----W----- 177
Db 5041 FFTKPLRNVDSVNGTCRLDCKTAGSLPMRVSWFKDGKEIAASDRYRIAPVEGTASLEII 5100
QY 178 -----N-----P-V-----S-D----- 182
Db 5101 RVDMDAGNFTCRATNSVGSKDSSGALI VQEPSPFVTKPGSKDVLPGSAVCLKSTFGST 5160
QY 183 -L-----L-----A-----SG----- 187
Db 5161 PLTIRWFGKNKELVSGGCVITKEALESLELYLVKTSQYTYTKVSNVAGGVECSANL 5220
QY 188 -----S-----GD----- 191
Db 5221 FVKEPATFVEKLEPSQLLKKGDATQOLACKVTGTPPIKITWFANDREIKESSKHRMFVES 5280
QY 192 TA--R----- 194
Db 5281 TAVRLTDVGIEDSGEYMCBAQNEAGSDHCSSIVIVKESPYFTKEPKPIEVLKEYDVMLL 5340
QY 195 -----I-W-N-----L-S-----E-----N----- 201
Db 5341 AEVAGTPPFEITWFKDNTILRSGRYKTFIQDHLVSLQILKFAADAGEYQCRVTNEVS 5400
QY 202 -----STS-----GST-----Q-----L-----V-L-----R 212
Db 5401 STCSARVTLREPPSPFKIESTSSLRGGTAAPQATLKGSLPITVTWLKDSDEITEDDNIR 5460
QY 213 -----H-----C-----I-R-----E----- 217

Db 5461 MTFENNVAISLYLSGIEVKHDKGYVCOAKNDAGIQRCALLSVKEPATITEEAVSIDVTOG 5520
QY 218 -----G-----GO----- 220
Db 5521 DPATLQVKSFGYKEITAKWFKDQBELTGLSKYKISVTDVTISILKIISTEKQDSGEYTFEV 5580
QY 221 --DV-----PS-NK-----D-VT-S-LD-----W----- 233
Db 5581 QNDVGRSSCKARINVLDLIIPPSFTKKLKMDSIKGSFIDLECI VAGSHFISIQWPKDDQ 5640
QY 234 -----N-----S-E-----GTL--AT--G-----S-----Y----- 245
Db 5641 EISASEKYKSFHDNTAFLEISQLEGTSGTYTCSATNAGHNQCSGHLTVKPEPPYFVEK 5700
QY 246 --D-----G-----F-----AR-IW-----T-----K-- 254
Db 5701 POSQDVPNTRVOLKALVGTAPMTIKWFDKNKELHSGAARSVWKDDTSTSLFLPAKAT 5760
QY 255 D-G-----N-----L-----AST----- 261
Db 5761 DSGTYICQLSNDVGTATSKATL FVKEPPQFIKKPSPVLVRNGQSTTFECQITGTPKIRV 5820
QY 262 --L-G-----Q-H-----K-----G----- 267
Db 5821 SWYLDGNEITAIOKHGISPIDGLATFQISGARVENSQTYVCEARNDAGTASCSELKVKE 5880
QY 268 -PIF--ALK-----W--N-----KK----- 277
Db 5881 PPTFIRELKPVEVYKSDVELECEVTGTPPPFVTLKNNREIRSSKKYTLTDRVSVFNLH 5940
QY 278 -----G-----N-----FI-----L-SA----- 284
Db 5941 ITKCDPSDGEYQCI VSNEGSCSTRVALKEPPSPFIKKIENTTTLVLSKSAFTQSVAG 6000
QY 285 -----GV----- 286
Db 6001 SPPISITWLKQDILDEDDNVYISFVDSVATLQIRSVNDHSGRYTQAKNESGVERCYA 6060
QY 287 ----- 286
Db 6061 FLLVQEPQIVKAKSVQVTEKDPMTLECVAGTPELKWLKDGKQIVPSRYFSMSFEN 6120
QY 287 --D-----K-----T-----T----- 290
Db 6121 NVASFRIQSVMKQDSQYTFKVENDEGSSSCDAYLRVLDQNI PPSFTKKLTMDKVLGSS 6180
QY 291 I-----I--W-D-----A-----H-----T----- 297
Db 6181 IHMECKVSGSLPISAQWFKDGKEISTSAKYRLVCHERSVSVLENNLEEDTANYTCKVSN 6240
QY 298 -----G-----E-----AK--QO-----P-----PF----- 306
Db 6241 VAGDDACSGILTVEPPSPFLVPGRQQAIPDSTVEFKAILKGTPPPKIKWFKDDVBLVSG 6300
QY 307 -----H----- 307
Db 6301 PKCFITGLESTSFNLNYSVDASKTQYTCVTNDVDSCTTMLLVTEPPKPKVLEASK 6360
QY 308 -----S-----A-----PA-----LD--V-----D 315
Db 6361 IVKAGDSSRLECKIAGSPEIRVVMFRNEHELPAISKYRMTFIDSVAVIQNNLSTEDSGD 6420
QY 316 ----- 316
Db 6421 FICEAQNPAAGTSCSTKVIKPEPPVFPSSPPPIVETLKNAEVLECELSGTTPPEVVMYKD 6480
QY 317 --Q-----S--N-----N-----T-----F-- 322
Db 6481 KQQLASSKKYKIAKNFHTSIHLNVDTSDIGEYHCKAQNVEGSDTCVCTVKLKEPPRFV 6540
QY 323 -----AS----- 324
Db 6541 SKLNSLTVVAGBPAELQASIEGAQPIFQWLKKEEVIRESINIRITTFVENVATLOFAKA 6600

QY 325 ---C---STD---M---C--- 330
Db 6601 EPANAGKYICQKNDGMEENWATLMVLEPAVIVEKAGPMVTVGETCTTLECKVAGTPEL 6660
QY 331 -I---H---V---C--- 334
Db 6661 SVEMWKDGLTSSQKHKFSFYNKISSLRILSVRQDAGTYTFQVQNNVKGSSCTAVVDV 6720
QY 335 ---K---LG--- 337
Db 6721 SDRVPPSFTRRLKNTGVLGASCILECKVAGSSPISVWPFHEKTIKIVSGAKYOTTFSDN 6780
QY 338 ---Q---D---R---P--- 341
Db 6781 VCTIQLNSLSDDMGNTCVANVAGSDECAVLTVOBPSPFVKEPPELVLPKNTFT 6840
QY 342 -I---K---T---F---Q---G--- 347
Db 6841 SVIRGTPPFKVNFRGARELVKGRCNYPEDTVAELELFNIDISQSGEYTCVVSNNAQ 6900
QY 348 ---H---T---N--- 350
Db 6901 ASCTTRLFVKEPAFLKRLSDHSVEPGKSHILESTYTGTLPISTVWTKDGNITTSKCN 6960
QY 351 ---EV---N---A---IK--- 356
Db 6961 IVTTEKTCILEILNSTKRDAGQYSCIEIENAGRDVCGALVSTLEPPYFVTELEPLEAAG 7020
QY 357 ---W---D---PT---G--- 361
Db 7021 DSVSLQOVAGTPEITVSWYKGDPTKLRPTPEYRTVFTNNVATLVFNKVNINDSGEYTC 7080
QY 362 -N---L---LA---SC---S--- 368
Db 7081 ENSIGTASSKTVFRIQBRQLPPSPARQLKDIEQTVGLPVLTLTCLRLNGSAPIQVCWYRDGV 7140
QY 369 ---DD---MTLKI--- 375
Db 7141 LLRDDENLQTSFVDNVATLKLQTLDSHQYSCSASNPLGTASSARLTAREPKKSPFF 7200
QY 376 ---WS--- 377
Db 7201 DIKPSIDVIAGESADPECHVTGAQPMRITWSKONKEIRPGNVTITCVGNTPHLRILKV 7260
QY 378 ---M---KQ--- 380
Db 7261 KGDSGYTQCOATNDVGKMCQAQLSVKPPKFVKKLEASKVAKQGESIQLECKISGSPE 7320
QY 381 ---D---N---C-VH--- 385
Db 7321 IKVSWFENDSELHESWKYNSFINSVALLTINEASAEKSDGYICEAHNGVGDASCSTALT 7380
QY 386 ---D---LQ---Q--- 389
Db 7381 VKAPPVFTQKSPVGVALKGSDVILQCEISGTPPEVWVKDRQVRNSKKFKITSKHFD 7440
QY 390 -H---NKE---I---Y---IK--- 398
Db 7441 SLHLNLEASDVGEYHCKATNEVGSDDTCSCSVKPEPPFRVKKLSDPTSLIGDAVELRAI 7500
QY 399 ---W---SP--- 401
Db 7501 VEGFQPSVWVKDRGEVIRESENTRISFDINTIATLQSGPEASNSGKYICQIKNDAGMR 7560
QY 402 ---TG-P---GT--- 406
Db 7561 ECSAVLTLEPARIIEKPEPMVTVTGNPFALCECVTGTPELSAKWFKDGRSLSDSKHHI 7620
QY 407 ---N---N---P---NA--- 411
Db 7621 TFINKVASLKIPCAEMSDKGLYSFEVKNVSGKNCTVSVHVSRIVPPSPFIRKLKDVNAI 7680

QY 412 ---NL-ML--- 415
Db 7681 LGASVLECRVSGSAPISVGVFODGNEIVSGPKQCSFSFENVCTLNLSLLEPDSGTGIYTC 7740
QY 416 ---A---SA---SF---DS---T---VR---L---W--- 427
Db 7741 VAANVAGSDSCSAVLTVQBPSPFEQTPDSVEVLPGMSLTFTSVIRGTPPPKVKWFKSRE 7800
QY 428 --- 427
Db 7801 LVPGESCNISLEDFVTELELFEVQPLESGDYSLVNDAGSACSTHLLFVKBPATFVKRL 7860
QY 428 -D---V---D---R--- 431
Db 7861 ADFSVETGSPVLEATYTGTPPISVSWIKDEYLIQSERCSTWTEKSTILELESTIED 7920
QY 432 ---G---IC---I---H---TL---T--- 439
Db 7921 YAQYSLIENAGODICEALVLEPPYFIEPLEHVEAVIGEPATLQCKVDGTFEIRISW 7980
QY 440 -K-H---Q---E--- 443
Db 7981 YKEHTKLRSAAYKMOFKNNVASLVINKVDHSDVGEYSCKADNSVGAVASSAVLVIKER 8040
QY 444 -P---V---YSVAF---S-P--- 452
Db 8041 LPFFPARKLKDVHETLGFVAFECRINGSPELQVSWYKGVLLKDDANLOTSPVHNATL 8100
QY 453 ---D---GRY-L-AS---G---S---FD---KC 465
Db 8101 QILQTOQSHIGQYNCASNPGLGTASSAKLILSEHEVPPFDLKPVSVDLALGESGTFKC 8160
QY 466 VH---I-W---NT---Q-T---G- 474
Db 8161 -HVTGAPAKITWAKDNREIRPGNYKMTLVENTATLTLVKVGKGAGQYTCVASNIAG 8219
QY 475 ---A---L--- 476
Db 8220 DSCSAHLGVQEPFRFIKLEPSRIVKQDEFTRYECKIGSGPEIKVLWYKDETEIQESSKF 8279
QY 477 ---V---H---S-Y---R--- 481
Db 8280 RMSFVDSVAVLEMHNLVSDSGDYTCEAHNAAGSASSSTLKVKEPPIFRKKPHPIETLK 8339
QY 482 GT---G---G---I---F---E--- 488
Db 8340 GADVHLECLQGTTPPHVSWYKDKRELRSKKYKIMSENFLLTSIHILNVDAADIGYQCK 8399
QY 489 ---V---C--- 490
Db 8400 ATNDVGSDDTCVGSIALKAPPRFVKKLSDDISTVVGVKEVQLOTTIEGAEPISVWVFKDKGEI 8459
QY 491 ---W---N---A-AG---D--- 496
Db 8460 VRESDNIMWISYENIATLQSRVPEANAGKYTCQIKNDAGMOECFATLSVLEPATIVEKP 8519
QY 497 ---KV-G---A---S---A-SD 504
Db 8520 ESIKVTGTCTLECTVAGTPELSTKWKFGDKELTSDNKKYSFFNKNVSGLKLIINVAPSD 8579
QY 505 -G-S---V---CV-L---D---L-RK 514
Db 8580 SGVYSEVQNPVKDSCSTASLQVSDTVPSPFTRK 8614
RESULT 12
AAP49011 PRELIMINARY; PRT; 7073 AA.
ID AAP49011
AC AAP49011:
DT 02-MAR-2004 (TREMELrel. 27, Created)
DT 02-MAR-2004 (TREMELrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMELrel. 27, Last annotation update)
DE Orfiab polyprotein.

QY 167 -----CHE-----S-----E----- 171
Db 2036 KEVIECDVKTEVGVNVLKPSDEGVKVTQELGHEDLMAAYVENTSITIKKNELSLALG 2095
QY 172 -----VF----- 173
Db 2096 LKTIATHGIAINSVPWSKILAYVKPLGQAAITTSNCAKRLAQRVNNMYPVFFLLFQ 2155
QY 174 -----I-----C-----A-W----- 177
Db 2156 LCTFTKSTNSRIRASLPTTIKNSVKSVAKCLDAGINYKSPKSLFTIAMWLLLSI 2215
QY 178 -----N----- 178
Db 2216 CLGSLICVTAAGVLLSNFGAPSYCNGVRELYLNSNVTTMDFCEGSPFPCISCLGLDSL 2275
QY 179 -----P-----DL-----LA-----S-----G----- 189
Db 2276 DSYPALETIQVTISSYKLDLTILGLAAEWLAYMLFTKFPYLLGLSALMQVFFGYFASHF 2335
QY 190 -----D-----ST-----A 193
Db 2336 ISNSLWMTIISIVQMAPSAMVRMYIFFASFYIWKSVYHMDGCTSSCTMCMCYKRNR 2395
QY 194 -----I-----WN-ISEN-ST--SGST----- 207
Db 2396 TRVECTIVNGMKRSFYVYANGRGFCCKTNWNL-NCDFCTGSTFTISDEVARDLSLQ 2453
QY 208 -----Q-L-----V-----L-----RH----- 213
Db 2454 FKRPINPTDOSSYIVDSVAVNGALHLYPDKAQKYERHPLSHFVNLDLNRANTKGS 2513
QY 214 -----C-I----- 215
Db 2514 PINVIVFGKSKDESASVYSQMLCQPIILLDQALVSDVGDSTEVSRMFDAYVD 2573
QY 216 -----R----- 216
Db 2574 TFSATFVPMEKULAVATAHSELAKGVALDGLVSTFVSAAQGVDTVDTKDVEICK 2633
QY 217 -----E-G-----G-----Q----- 220
Db 2634 LSHSDLEVTGDSNNFMTYNKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVK 2693
QY 221 D----- 221
Db 2694 DYNLSBQLRQIRSAAKNNIPRLTCAITRQVNVNITTKISLKGKIVSTCFKMLKA 2753
QY 222 -----V-P-----S-----N-----K-----D-VT-----S-----LD- 232
Db 2754 TLLCVLAALVCYIYMPVHTLSIHGTYNEIIGYKAIQDGVTRDIIISTDDCFANKHAGFDA 2813
QY 233 W-----N-S-----E-----GT----- 238
Db 2814 WFSQSGSYKNDKCPVVAALITREIGFIVPGLPGTVLRAINGDFLHFLPRVFSAVGNIC 2873
QY 239 -----L-----LA-T-----G-----SY----- 245
Db 2874 YTPSKLIEYDFATSACVLAECTIFPKDAMKPVPCYDNTLLEGSISYSELRPDTRYVL 2933
QY 246 -----DG-----F-A-----R-----I-----W----- 252
Db 2934 MDGSIIOFPNTYLEGSVRVVTTFDAEYCRHGTCERSEVGICLSTSGRWVLNNEHYRALSG 2993
QY 253 -----TK-----D----- 255
Db 2994 VFCGVDMNLIANITFPLVQPVGALDVSASVAVAGGIITAILVTCAAYFMKFRFVGEYNH 3053
QY 256 -----G-----N-----LA-----S----- 260
Db 3054 VVAANALLFLMSFTILCLVPAYSLPGVSYVLYLTFTYFTNDVSFLHLQWFAFPSPIV 3113

QY 261 -----T-----L-----G----- 263
Db 3114 PFWITAIYVFCISLKHCHWFFNNYLRKRVFMGVFTSTFEBAALCTFLLNKEMYLKRSE 3173
QY 264 -----Q-----H-K-----G-----P----- 268
Db 3174 TLLPLTQYRNRYLALYNKYKYSFSGALDITTSYREAAACHLAKALNDFSNSGADVLYQPOTS 3233
QY 269 I-----F-AL--K-----W----- 274
Db 3234 ITSAVLQSGFRKMAFPGSGKVEGCMQVTCGTTTLNGLWLDITVYCPRHVICTAEDMLNPN 3293
QY 275 -----N-----K----- 276
Db 3294 YEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKRVRIQPGQTS 3353
QY 277 -----KG--N-----F-I-L--S-----A 284
Db 3354 VLACYNGSPGVYQCAMPNHTIKGSFLNGSCGVGFNIDYDCVFCYMHMELPTGVHA 3413
QY 285 G-----VDK-----TTI-----I--W-----D----- 294
Db 3414 GTDLBGKFGPFDROTAQAAGTDTITLNLVLAWLYAAVINGDRMFLNFTTLNDPNLV 3473
QY 295 A-----H-----TG-----E----- 299
Db 3474 AMKYNVEPLTQDHDVILGPLSAQTGLAVLDMCAALKELLQNGMNGRTILGSTILEDEFTP 3533
QY 300 -----AK-----Q-Q-----P-PP-- 306
Db 3534 FQVVRQSGVTFQGRFKIVKGTTHHMLLTFLTLLILVQSTQWSLFFPVYENAFLEFTL 3593
QY 307 -----H-----S--A--PA----- 311
Db 3594 GIMATACAMLLVKGHAFCLFLPLSLATVAYFNVMPASVWVRIMTWELADTSLG 3653
QY 312 --L-D-----V-D--W-----Q----- 317
Db 3654 YRLKDCVMVASALVLLILMTARTVYDDAARRVMTLNVITLVVYVYGNALDOAISMWAL 3713
QY 318 -----SN-----NT----- 321
Db 3714 VISVTSNYSGVTTIMFLARAIVFCVEYVYPLFITGNTLQCIIMLVYCFGYCCCYFGL 3773
QY 322 FASC-----ST-----D----- 328
Db 3774 F--CLLNRYFRLTLGVYDLVSTQEFYMNVSQGLLPKSSIDAFKLNILKLGIGGKPCIK 3831
QY 329 -----M-----CI-H----- 332
Db 3832 VATVQSKMSDVKCTSUVLLSVLQQLRVESSSKLWAOQVOLHNDILLAKDTTEAFKMWSL 3891
QY 333 -----V--CK-----L-----G-- 337
Db 3892 LSVLLSMOAGVDTNRLCEEMLDNRATLOAIASEFSLPSYAAAYATAQAEYEQAVANGDSE 3951
QY 338 -----Q----- 338
Db 3952 VULKUKKSLINVAKSEFDRDAAMQKLEKXADQAMTQYKQARSEDKRAKVTSAMQTMFL 4011
QY 339 -----D-----R--P--I-----K-----TF----- 345
Db 4012 TMLRKLNDALNNIINNARDGCVPLNIIPLTTAAKLMVVVDPDYGTYKNTCDGNTFTYASA 4071
QY 346 -----Q--CH-----T-----NE--V----- 352
Db 4072 LWBIQVVDADSKHVQLSEINMDNSNLAWPLIVTALRANSVKLQNNELSPVALQMSC 4131
QY 353 -----NA-----I-----KW-----D-----P-- 359
Db 4132 AAGTTQACTDDNALAYNNKGRFVALLSDHQDLKWARFPKSDGTGTIYTELEPPCR 4191
QY 360 -----T-----G--NL-----LA-----S-CS-- 368

RA Nikaido I., Osato N., Saito R., Suzuki H., Yamana I., Kiyosawa H.,
RA Yagi K., Tonaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schirali L.M., Kanepin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusci V., Ciothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sakaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -|- FUNCTION: Potential cell surface proteins that bind and
CC internalize ligands in the process of receptor-mediated
CC endocytosis.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -|- SIMILARITY: Belongs to the LDLR family.
CC -|- SIMILARITY: Contains 17 EGF-like domains.
CC -|- SIMILARITY: Contains 32 LDL-receptor class A domains.
CC -|- SIMILARITY: Contains 25 LDL-receptor class B domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF270884; AAF81558.1; -;
DR EMBL; AK035795; BAC29188.1; -;
DR EMBL; AK034122; BAC28594.1; -;
DR HSSP; Q07954; 1CR8.
DR MGD; MGI-2151136; Lrp1b.
DR InterPro; IPR011044; Amine_DH_B_like.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR Pfam; PF00008; EGF; 15.
DR Pfam; PF00057; Ldl_recept_a; 32.
DR Pfam; PF00058; Ldl_recept_b; 33.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00179; EGF_CA; 4.
DR SMART; SM00192; LDLA; 32.
DR SMART; SM00135; LY; 36.
DR PROSITE; PS00010; ASX_HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS50026; EGF_3; 9.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS01209; LDLRA_1; 27.
DR PROSITE; PS50068; LDLRA_2; 32.
KW Calcium-binding; EGF-like domain; Endocytosis; Glycoprotein; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 20 Potential.

CHAIN	21	4599	Low-density lipoprotein receptor-related protein 1B.
FT DOMAIN	21	4444	Extracellular (Potential).
FT TRANSMEM	4445	4467	Potential.
FT DOMAIN	4468	4599	Cytoplasmic (Potential).
FT DOMAIN	31	70	LDL-receptor class A 1.
FT DOMAIN	76	114	LDL-receptor class A 2.
FT DOMAIN	120	154	EGF-like 1.
FT DOMAIN	155	194	EGF-like 2, calcium-binding (Potential).
FT DOMAIN	195	336	LDL-receptor class B 1.
FT DOMAIN	338	380	LDL-receptor class B 2.
FT DOMAIN	382	424	LDL-receptor class B 3.
FT DOMAIN	568	609	LDL-receptor class B 4.
FT DOMAIN	611	655	LDL-receptor class B 5.
FT DOMAIN	657	705	LDL-receptor class B 6.
FT DOMAIN	707	754	LDL-receptor class B 7.
FT DOMAIN	794	834	EGF-like 3.
FT DOMAIN	844	882	LDL-receptor class A 3.
FT DOMAIN	885	923	LDL-receptor class A 4.
FT DOMAIN	926	963	LDL-receptor class A 5.
FT DOMAIN	966	1003	LDL-receptor class A 6.
FT DOMAIN	1005	1043	LDL-receptor class A 7.
FT DOMAIN	1052	1089	LDL-receptor class A 8.
FT DOMAIN	1094	1132	LDL-receptor class A 9.
FT DOMAIN	1135	1174	LDL-receptor class A 10.
FT DOMAIN	1172	1213	EGF-like 4.
FT DOMAIN	1300	1345	LDL-receptor class B 8.
FT DOMAIN	1347	1388	LDL-receptor class B 9.
FT DOMAIN	1390	1435	LDL-receptor class B 10.
FT DOMAIN	1437	1480	LDL-receptor class B 11.
FT DOMAIN	1481	1521	LDL-receptor class B 12.
FT DOMAIN	1527	1570	EGF-like 5.
FT DOMAIN	1618	1659	LDL-receptor class B 13.
FT DOMAIN	1661	1703	LDL-receptor class B 14.
FT DOMAIN	1705	1743	LDL-receptor class B 15.
FT DOMAIN	1745	1788	LDL-receptor class B 16.
FT DOMAIN	1834	1875	EGF-like 6.
FT DOMAIN	1922	1963	LDL-receptor class B 17.
FT DOMAIN	1965	2006	LDL-receptor class B 18.
FT DOMAIN	2008	2050	LDL-receptor class B 19.
FT DOMAIN	2052	2094	LDL-receptor class B 20.
FT DOMAIN	2143	2183	EGF-like 7.
FT DOMAIN	2239	2278	LDL-receptor class B 21.
FT DOMAIN	2281	2328	LDL-receptor class B 22.
FT DOMAIN	2330	2373	LDL-receptor class B 23.
FT DOMAIN	2375	2416	LDL-receptor class B 24.
FT DOMAIN	2417	2458	LDL-receptor class B 25.
FT DOMAIN	2509	2548	LDL-receptor class A 11.
FT DOMAIN	2551	2587	LDL-receptor class A 12.
FT DOMAIN	2590	2626	LDL-receptor class A 13.
FT DOMAIN	2629	2675	LDL-receptor class A 14.
FT DOMAIN	2681	2717	LDL-receptor class A 15.
FT DOMAIN	2719	2757	LDL-receptor class A 16.
FT DOMAIN	2760	2800	LDL-receptor class A 17.
FT DOMAIN	2804	2841	LDL-receptor class A 18.
FT DOMAIN	2844	2885	LDL-receptor class A 19.
FT DOMAIN	2890	2926	LDL-receptor class A 20.
FT DOMAIN	2926	2967	EGF-like 8.
FT DOMAIN	2968	3008	EGF-like 9.
FT DOMAIN	3055	3097	LDL-receptor class B 26.
FT DOMAIN	3099	3140	LDL-receptor class B 27.
FT DOMAIN	3142	3184	LDL-receptor class B 28.
FT DOMAIN	3186	3223	LDL-receptor class B 29.
FT DOMAIN	3225	3266	LDL-receptor class B 30.
FT DOMAIN	3274	3314	EGF-like 10.
FT DOMAIN	3316	3353	LDL-receptor class A 21.
FT DOMAIN	3356	3392	LDL-receptor class A 22.
FT DOMAIN	3395	3432	LDL-receptor class A 23.
FT DOMAIN	3435	3472	LDL-receptor class A 24.
FT DOMAIN	3475	3511	LDL-receptor class A 25.
FT DOMAIN	3514	3550	LDL-receptor class A 26.
FT DOMAIN	3552	3588	LDL-receptor class A 27.
FT DOMAIN	3593	3629	LDL-receptor class A 28.

FT	DOMAIN	3631	3668	LDL-receptor class A 29.	Db	538	MDLNTKIADCEMIP	PIENLVNPRALDPFAEANYIYFADTTTFLIGRQKIDGTERETILKOD	597
FT	DOMAIN	3673	3711	LDL-receptor class A 30.	Qy	85	-D----	AV-----M-----P----	89
FT	DOMAIN	3714	3752	LDL-receptor class A 31.	Db	598	LDNVEGIAVDWIGNLY	WTNDGHRKTINVARLEKASQSRKTLLEGGMSPRAIIVDPVNG	657
FT	DOMAIN	3761	3797	LDL-receptor class A 32.	Qy	90	-----D-V-----V-Q-----	-----Q-----	93
FT	DOMAIN	3801	3843	EGF-like 11.	Db	658	MMWTWDKEDKIDDS	VGRIEKAMWDGVRQVFTVSKMLPENGTLDFHTSTLYWCDAIYD	717
FT	DOMAIN	3844	3881	EGF-like 12.	Qy	94	-----T-R-----	-----Q-----	96
FT	DOMAIN	3981	4037	LDL-receptor class B 31.	Db	718	HIEKVFNGTHRKVVY	SGKELNHPFGLSHHGNYVFTWDMYNGSIFQLDLMTNEVTLRHE	777
FT	DOMAIN	4039	4081	LDL-receptor class B 32.	Qy	97	-----QAV--R-----D-K-----LA-----	-----LA-----	104
FT	DOMAIN	4083	4126	LDL-receptor class B 33.	Db	778	RAPLFGLOIYDRKQ	QOGDMCRINNGCGTLCIAIPAGRVACADNQLLDENGTTCTFNP	837
FT	DOMAIN	4213	4249	EGF-like 13.	Qy	105	-----Q-----	-----Q-----	107
FT	DOMAIN	4249	4290	EGF-like 14.	Db	838	EEIRPHICKPGEFRCK	NKHCIQARWKCDGDDCLDGSDESDSVTCFNHSCPDQDFKQNNR	897
FT	DOMAIN	4286	4326	EGF-like 15.	Qy	108	-----A-----A-----AA-----AA-----AA-----	-----AA-----	116
FT	DOMAIN	4321	4363	EGF-like 16.	Db	898	CIPKRWLCGDANDCG	SNEDESNQTCCTARTCAQDFSCGNRCIPTAWLCDREDDCGQTD	957
FT	DOMAIN	4392	4430	EGF-like 17.	Qy	117	-AAS-----Q-----	-----Q-----	121
FT	DISULFID	32	45	By similarity.	Db	958	EVASCEFPTEBLTQ	FIKSGRCISNKHCHDTHDDCGDRSDEVGCVHSCLDQDFRCSGR	1017
FT	DISULFID	39	58	By similarity.	Qy	122	--GS-A--KN--G--E--N-T--A--N--	-----N-----	134
FT	DISULFID	52	69	By similarity.	Db	1018	CIPGHWACDGDND	CGDFSDETHINCKEARSAGCIGNEFCQRCPDGNCIPDLWRCDGK	1077
FT	DISULFID	77	90	By similarity.	Qy	135	-----E--NGA-----H-----T--I-----	-----T--I-----	141
FT	DISULFID	84	103	By similarity.	Db	1078	DCEDSGDEKGCNGT	IRLCHDKTKFSCRSTGCRCINNAWCDGVDCEQDEEDCDSFLCG	1137
FT	DISULFID	97	113	By similarity.	Qy	142	-----AN-----	-----NH-----	145
FT	DISULFID	159	169	By similarity.	Db	1138	PPKYPACANDTSVCL	QPEKLCNGRKDCPDGSDGDLCECSLNGGCSNHCSSVVGRIVC	1197
FT	DISULFID	165	178	By similarity.	Qy	146	-----T-----D-----MM-----E-----	-----E-----	152
FT	DISULFID	180	193	By similarity.	Db	1198	SCPEGHQKNDRTCE	IVDYCASHLRCSQVCEQKHVKSCVEGHALGTGDSCTSVDS	1257
FT	DISULFID	798	809	By similarity.	Qy	153	-----GD-----	-----GD-----	154
FT	DISULFID	805	818	By similarity.	Db	1258	FEAFIIFSIIRHEIR	IDLHKGYSLLVPLGRNTIALDFHFNQSLLYWTDVVEDRIYRGKL	1317
FT	DISULFID	820	833	By similarity.	Qy	155	-----VE-----	-----I-----	157
FT	DISULFID	845	857	By similarity.	Db	1318	SSGGVSAIEVVVEH	GLATPEGLTVDWNIAGNIYWDNSLDQIEVSKLDGSLRATLIAGAM	1377
FT	DISULFID	852	870	By similarity.	Qy	158	-----P-----P-----N-----K-----	-----K-----	161
FT	DISULFID	864	881	By similarity.	Db	1378	EHPRALDPRYGILF	PTWDNDANPRIESASMSGAGRTIYKDMKTGAMPNGLTVDHFER	1437
FT	DISULFID	886	898	By similarity.	Qy	162	-----AV-----	-----VLRGHE-----SEV-----	172
FT	DISULFID	886	898	By similarity.	Db	1438	RIVWTDARSDAIYS	AFDGTNNMIEIRGHEYLSPFAVSLYGVYVWTDWRTNTLAKANK	1497
FT	DISULFID	886	898	By similarity.	Qy	173	-----F-----	-----F-----	176
FT	DISULFID	886	898	By similarity.	Db	1498	WTQNVSVIQTSAOF	LDIYHPSRQPOAPNPCAANEGRGPCSHLCLINHNRSACACP	1557
FT	DISULFID	886	898	By similarity.	Qy	177	-----W-----	-----W-----	182
FT	DISULFID	886	898	By similarity.	Db	1558	HLMLSSDKKTCYEM	KKFLLYARRSEIRGVDDNPNVNFITAFVTPDIDDAVIDDASE	1617
FT	DISULFID	886	898	By similarity.	Qy	183	-----LL-----	-----A-----SG-----S--G--D-----S--T-----	192
FT	DISULFID	886	898	By similarity.	Db	1618	ERLYWTDIKTQIT	ITRAFINGTGLTISRDIQSIIRGLAVDWSRNLVWISSEFETQINV	1677

Query Match 71.3%; Score 2610.5; DB 1; Length 4599;

Best Local Similarity 9.5%; Pred. No. 1.2e-41;

Matches 434; Conservative 56; Mismatches 16; Indels 4085; Gaps 365;

Qy	1	MS-----I-----S-----	4
Db	1	MSQLLAILTLGSLPNAEVLIVGANOQHLCPGDFLCHDHVTCVSQSWLCDGDPDCPD	60
Qy	5	-SD-----EV-----N-----F-----	10
Db	61	QSDSLDTPCEVEIKCPLNHIAHGSSACVHLSKLCNGVVDGPDGDFEGHGCQELLPS	120
Qy	11	--L-----V-----Y-----R-----Y-L-Q-----	17
Db	121	QQLNCOFKAMVRNATFCYCEDGFEVAEDGRSCKDQDECSIYICISQTKNTYGSVAC	180
Qy	18	-ESGF-----S-----H-----SAF-T--F--GTX-----S-----H-----	33
Db	181	VE-GYIMQSDNRSCVKHKEPTKAPMLLISLETIELFYINGSKMTLSSANRNEHTLD	239
Qy	34	-----I-S-QS-----NI-----	39
Db	240	FIYSEMICWIESRESSNLKCGOITKAGRLTDRIINLSQSFQNEQMAFDWLTRNIYF	299
Qy	40	-----NG-----A-L-----VP-----	45
Db	300	VHVSDFRIFVCNENGSCVTLIBSELNPKAIAADPIAGKLFPTDGNVPKIERCDLGM	359
Qy	46	-----P-A-AL-----I-----SIIQKGLQ-----Y--59	
Db	360	NRTRIVYSKAEQPSALALDNLRLVYVDLYDVGVDYQGNRHTIVQ-GRQVKHLYG	418
Qy	60	--V-E-----AEVS-----IN-----	68
Db	419	ITVFEDLYA-TSSDNFNIRNFRNGTDIHSIKMESARGINTYQKRTQPTVRSACEV	477
Qy	69	D-----G-----TL-----F-----DGR-----	76
Db	478	DAYCMGGGSHICLLSSSYKTRCTCRCTGFNMGSDGRSCKRPKNELFLFYKGRPGIVRG	537
Qy	77	-----PIE-----SL-----S-LI-----	84

QY 193 AR-----I-----W-----N-----197
Db 1678 ARLOGLSKTSIIHGIDXPQCLAAHPVRGKLYWTGNTINMANMDGNSKILFQKQKEPVG 1737
QY 198 LS-----S-----T-----SG-----S-----T-----207
Db 1738 LSIYVENKLYWISSNGTINRCNLNLEVIEMKEELTKATALTIMDKLWADONL 1797
QY 208 OL-----VLR-----H-----213
Db 1798 AOLGTCKNRDGRNPSILRNKTSIGVVMKVDYDKEAQOGNSQVNNNGCSQLCLPTSETTR 1857
QY 214 -----C-----IR-----216
Db 1858 TCMCTVGYLQKRNMSQGIIESFLMYSVHEGIRGIPLEPRDKVDALMPISGAAPAVGIDF 1917
QY 217 -----EG-----218
Db 1918 HAENDTIYTDMLNKLISRAKRDQTKWEDVVTNGLGRVEGIAVDWIAAGNIYWTDDHGFNLI 1977
QY 219 -----GO-----D-----V 222
Db 1978 EVARLNGSFRYVIIISQGLDOPRSIAVHPEKGFLEFWTEWQVPCITGKARLDGSEKVMIVSV 2037
QY 223 ---P---S---NK---D---V---T---S---LD---232
Db 2038 GITWPNGISDIYENKLYWCARDSDKIERIDLTGANREVLVSGSNVDLFSVAVFGAIYI 2097
QY 233 W---N---S-----E---GT-----238
Db 2098 WSDRAHANGSVRRGHKNKDATETVMTRTGLGVNLKEIKIPNRVREKGTNVCAKENGCGQQL 2157
QY 239 -----LLA---T-----G-----S---Y---245
Db 2158 CLYRGNRRTCACAHGLYLAGVTCRLRHEGYLLSGRTILKSILHSDETNLSVPRPYEN 2217
QY 246 -----D---G---F---A-----R---I-----251
Db 2218 PNYFNIIALAFDYNORREGNRIIFYSDAHFNGIQLIKDNWEDRQVIVENVGSVEGLAYH 2277
QY 252 -----W-----TK---D---G-----NL-----258
Db 2278 RAWDTLYWTSSTSSITRHTVQTRPCAIDREAVITMSDDHPHVLADCEQNLFWFTNW 2337
QY 259 -----A---ST-----LG-----Q---264
Db 2338 NEQHPSIMRATLTGKNARVVVSTDLTPNGLTIDHRAEKLYFSDGSLGKIERCEYDGSQR 2397
QY 265 H---K---GP-----IF-ALKW-----NK-----K-----277
Db 2398 HVIVKSGGTFLSLAVYDSYIFWS-DWGRRAILRSNKYTGGETKILRSDIPHOPMGIIV 2456
QY 278 -----G-----280
Db 2457 ANDTNSCELSPCALLGGCHDLCLITPDGRVNCSCRGDRVLLANNRCVTKNSCNIIYSEP 2516
QY 281 -----I-----L-----S---AG-----V-----286
Db 2517 ECGNGDCVDYVLTCDGIPHCKDKSDEKLLYCENRSCRSRGFKPCYNRRRCVPHGKLCBDGTD 2576
QY 287 -----D---K---T-----T---I-----291
Db 2577 CGDSSDELDCKVSTCSTVEFRCADGTICIPRSARCNQNMDCSDASDEKGCNNTDCTHYKL 2636
QY 292 -----I-----W-----D-----AH-----296
Db 2637 GVXSTGFI RCNSTSLCVLPWSWICDGSNDGCDYSDLKPCPVONKHCBENYFGPCSGRCIL 2696
QY 297 ---T---G---E-----A-----K-----301
Db 2697 NTWVCDGQKDCBDGLDLHCDSSCSMNQFACSVKCKISKHWICDGBDDCGDSLDESISIC 2756

QY 302 -----Q-----Q-----F-----304
Db 2757 GAVTCAADMPSCOGSHACVPQHWLDCGERDCPDGSDLSAGCAPNNTCDENAFMCHNKV 2816
QY 305 ---P---F---H-----S-----AP-----ALD---VD---WQ-----S 318
Db 2817 CIPKQFVCHDDDDCGDGSDEFLQCGYRQCGPEFRCA-DGRCLVNTLWQCDGDFDCPDSS 2875
QY 319 ---N-----NT---FASC---STD-MC-I-----HV---C---334
Db 2876 DEAPINPRCSRABHSNSSFMM-CKNGRCIPS-DGLCDIRDDCGDSDETNCHINECLSK 2933
QY 335 KL-G---OD-R---PI-----KT-----P---Q---G-----347
Db 2934 KISGQCDQDLPSYKCKWPGFQLKDGKTCVDIDECSSGPPCQOCINTYGTKYCHC 2993
QY 348 -----H-----T-----N-----350
Db 2994 AEGYETOPDNPNGCRSLSDPEPLIADQHEIRKISTDGSNYTLLKQGLNNVIALDFDYR 3053
QY 351 -E-----V-NA-----I-K---W-D-----358
Db 3054 EEFIYIDSSRPNGSRINRMCLNGSDIKVHNTAVPNALAVDWIGKLYWSDTEKRIIEV 3113
QY 359 -----PT-----GNLL-ASC-----367
Db 3114 SKLNGLYPTVLYSKRLKFPRLDLSLDPAGNLYWIDCEYPHIGRVGMDGTNQSVIETKI 3173
QY 368 S-----D-----M-----TL-K-I-W-----376
Db 3174 SRPMALTIDVNHRLYADENHIEFSNMDGSHRHKVPNQDIPGVIALTLFEDYIYWDGK 3233
QY 377 ---S---MK-----Q-D-----N-----C-----383
Db 3234 TKLSLRVHTKSGADRLSLINSHAITDIQVHYSYRQPDVSKHLCTVNNGGCSHLCLLPG 3293
QY 384 -VH-----D---L-----387
Db 3294 KHTCACTNFYLAADNRCTCLSNCTASQPRCKTDKIPFWKCDTVDDCGDSDDEDDCP 3353
QY 388 ---Q-----Q-----H-----NKE-I---394
Db 3354 EFKQCPGRFCQGLCALPAFICDGENDCGNSDELNCDTHVCLAQFCKTKNKKCIPVN 3413
QY 395 -----Y-----T---I-K-W---SP-----401
Db 3414 LRNGQDDCGDEDEKDCPENSCSPDYFOCKTTKHCISKLWVCDDEDDPCADASDEANDCK 3473
QY 402 -T-GP-----G-----T-N---N---P-----409
Db 3474 KTCGPHEFOCKNNNCIPDHWRCNQDNCSDNDEDNCKPQTCTLKDFLCSNGDCVSSRW 3533
QY 410 -----N-----A-----N-----412
Db 3534 CDGEFDCADGSDXKNCETSCSKDQFQCSNGQCLSAKWKCDGHEDCKYGEKNCPEAPV 3593
QY 413 -----LMLA---SNS-----F-----420
Db 3594 CSSESVMCASGGCLSASLKNCEPDCVDSDEMDCVIECKEDQFOCKNKAYCIPIRWLCD 3653
QY 421 -----D-S---T---V-R-----L-W-----D---V---429
Db 3654 GIYDCVDSDEFTCGRGGSI CRDDEFLNNLSLKLHFWVCDGDDCGDSDENAPMCVKF 3713
QY 430 -----DR-----GI-----433
Db 3714 LCPPTPRYCRNDRICLOLEKICNGINDCGDSDSEHCSGLSKSPCKKDEFTCSNRN 3773
QY 434 CI-----HT-----437
Db 3774 CIPMELQCSLDDCGDGSDEOGLKTPTEHTCENNGNPGCDDAYCNQIKTSVFCRCKPGF 3833
QY 438 -----L-T-----K-----440

Db 3834 QRNMKGRECADNELLFGICSHCLNTRGYSKVCQDNQFQEKNSCIAKSGEDQALYIA 3893
QY 441 -----HQ-----E-----P--V-Y-----446
Db 3894 NDTDLGFFVFFNYSGHQIISHVHNSRITGMDVHYQRNVIVWSTQFNGGIFYKMDA 3953
QY 447 -----S-----VA-----FS-----451
Db 3954 REKQANSGLICPEFKRPDIADVAGNVVTDHSMHNFYSYTHWTSLSRYNSVQGL 4013
QY 452 -----P-----DG-----454
Db 4014 NGPNCRTLLTNMAGEPYAIAVNPGRGMVYTWVIGDHSIHBEAAMDGTLRVVLQKNLQRP 4073
QY 455 -----R-Y-----LA--S-----GS-----461
Db 4074 TGLTVDFHGRITWADFELSIIGSVLYDGSPPVSVSSKOGLLPHRPHIDVFEDYIYAGP 4133
QY 462 -----F-----DK-----464
Db 4134 KNGIFRVQKFGHGSVEVLALGVDKTKSILVSHRYKQLNLPNCLDLSCDFLCLNPSGAT 4193
QY 465 CV-----H-----I-----W-----N--470
Db 4194 CICPEGYMMNGTCHDDSLDSDCKLTCENGRCILNEKGLRCHWCWPSYSGRCRVCNHC 4253
QY 471 T--Q-----T-G-----AL-----V-----H--S-----479
Db 4254 SNYQNGGTICPSTLGRPTICALGFTGPNCGKAVCEDSCHNGSCVVTAGNPYCHCQA 4313
QY 480 Y--R-----G-----T-----G-----485
Db 4314 DYTGDRCQYVCHHYCVNSSECTIGNDGSVEVCPTRYEGPKCIDKVCRCGHGCLINK 4373
QY 486 --IF-----E-----VC---W-----491
Db 4374 DNEDIFCNCTNGKIASSCOLGDCYNGGTGCTQDPETSPVCVCSNWSGTQCRPAPKS 4433
QY 492 -----491
Db 4434 SKSEHISTRISAIIVPLVLLVTLTVLVLGVCKKRRKTKTTRROPINGINVEIGNP 4493
QY 492 --NA-----AG-----D-K--VG---A-----SA-----SD-----G 505
Db 4494 SYNMYEVDHDSGGLLEPFSMDPVKSRVIGGSSAPKLPHTAPPIYLNLDLKGPLITFG 4553
QY 506 --S--V-----C-----VLDLRK 514
Db 4554 PTNYSNPVYAKLYMDGQNCNRSLSAV-DERK 4583

RESULT 14

Q8NJX1 PRELIMINARY; PRT: 20925 AA.
ID Q8NJX1
AC Q8NJX1
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Nonribosomal peptide synthetase.
GN Name=pep;
OS Trichoderma virens.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocrea; Hypocrea.
OX NCBI_TaxID=29875;
RN [1]
RX MEDLINE=22037864; PubMed=11909873;
RA Wiest A., Grzegorski D., Xu B.W., Goulard C., Rebuffat S.,
RA Ebbole D.J., Bodo B., Kenerley C.;
RT "Identification of peptaibols from Trichoderma virens and cloning of a
RT peptaibol synthetase."
RL J. Biol. Chem. 277:20862-20868(2002).

[2]
RN SEQUENCE FROM N.A.
RP Wiest A.E., Grzegorski D., Xu B.-W., Zhou W., Ebbole D.J.,
RA Kenerley C.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC EMBL: AF469045; AAM78457.1; --
DR HSP: O30409; 1DNV.
DR GO: GO:0048037; F.cofactor binding; IEA.
DR GO: GO:0016740; P:transferase activity; IEA.
DR GO: GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR010071; AA adenyl_dom.
DR InterPro: IPR009081; ACP like.
DR InterPro: IPR001227; Ac.Transferase.
DR InterPro: IPR008873; AMP-bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR000794; Ketoacyl_synth.
DR InterPro: IPR006162; P:antn S.
DR InterPro: IPR006163; P:bind.
DR InterPro: IPR010080; Thioester redct.
DR Pfam: PF00698; Acyl.transf_1; 1.
DR Pfam: PF00501; AMP-binding; 18.
DR Pfam: PF00668; Condensation; 27.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF00550; PP-binding; 19.
DR PRINTS; PR00154; AMPBINDING.
DR TIGRFAMS; TIGR01733; AA-adenyl-dom; 17.
DR TIGRFAMS; TIGR01746; Thioester-redct; 1.
DR PROSITE; PS00455; ACP DOMAIN; 19.
DR PROSITE; PS00455; AMP BINDING; 18.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN 11.
SQ SEQUENCE 20925 AA; 2320297 MW; C6603A269D56FE11 CRC64;

Query Match 71.3%; Score 2607.7; DB 2; Length 20925;
Best Local Similarity 8.1%; Pred. No. 2.2e-40;
Matches 476; Conservative 32; Mismatches 3; Indels 5343; Gaps 395;

QY 1 M-----SI-----SS-D--EV-----N-----9
Db 6914 MIVPSSVPCBGLTSIMVETIIELEQLSSRYDAFQEIILPKAEPNAAVVLFTSGTGK 6973
QY 10 -----F--L--VV-----R--Y-----15
Db 6974 PKGLMEHSAFATSLTCHGILNLSPASRVQFSNIFDGLGEIFTTLTSGTVCVPS 7033
QY 16 --LQ-----E-S-----19
Db 7034 DERLQKAPSMREARVNTAMLTPTSFVTRTAPQVPSLRLLVLGGEPSSKDLLETWCRLR 7093
QY 20 -----G-----F--S-H-----SA-----25
Db 7094 LVNGYGPABACNAYATTHDFKPTDSPHTIGRGFNSACWIVDPTDYNKLTPIGCIIGLIQ 7153
QY 26 -----F-T-----F-----28
Db 7154 NALARGYINDARTKNSFITNVDCLPKSIISGPHRYLTGDLVRYTPDGOLEYLGRKDTQ 7213
QY 29 -----G-I-----KS-----H-----I--S-Q-----S-NI---39
Db 7214 VKLGRGLEGEIEYHVKKSLANIEHVAVDVAHREGTDTLIAFVSPKMKWATTSNILL 7273
QY 40 N-----G-----A-----42
Db 7274 NDDLRLVALATIMEHLKMSLPGYMVPSTILPVRMPFITSMKVDKRLTAMAAALSLEIT 7333
QY 43 --LV-----P--P-----46
Db 7334 SPSLVKRDYIASTPWEKMLANLWAOVLKISSIEDICKNDSFLQIGGDSSTSHLSLAQ 7393
QY 47 -----A-----AL-----I-----S 51

Db 7394 SGISLTVASIFADSKLSTMAISAGDGKIEPIROVPEFMAPTRDLDTLLDVRKACGLSS 7453
QY 52 --I--I--OKGL--Q-Y--V--EA--E-- 63
Db 7454 TAIVEDIYPTRFQEGMALAVKQPGSHIAKQYRLSQNVVDIARFKAAWMTWELCSNLR 7513
QY 64 --V--SI--N--E-- 68
Db 7514 TRIVLSDASIQVLKNVRWEDTSNMTLSYLETNQNSHMYGSALSRRHALIEADGNH 7573
QY 69 --DG--TLF--D--G-- 75
Db 7574 YFVMSIHAVFDGWTTRILLTFLSAVMDGEITPLNPTRFINVTYRSLDADAIAKIYWGEO 7633
QY 76 --R--P--IE--S-I--S-L-- 83
Db 7634 LLDKRALFPTISNGVPRNKSNTRVFEKAIELEFRFKQTSITWASILRATWSIVLAQHCD 7693
QY 84 --I--D-- 85
Db 7694 TNDVTFGTALSGROAPVPGITEKAGPVVATVPVRVRLDKLASVKFLQSQNOASEMIPF 7753
QY 86 --A--V--M--P--D-- 90
Db 7754 EQFLQNIAXLNVKEACDFSSLLVTQPMKLLDGNRKPILHMSATTEDSEDHMQNY 7813
QY 91 --VVQ--T-- 94
Db 7814 FSYPLVIOGHVYEDSVNLVLIYDADILPEOQLLAHAHQFHVAAQLVAEDSGTMPLGNVS 7873
QY 95 --R--R--Q--A--Y-- 100
Db 7874 VSGSWDIEYALRQNNETPLRIESCFTFVIEQOIIIRPESAVHGDGNTFYQGLDQANR 7933
QY 101 --D--K--LA--Q--Q-- 107
Db 7934 LANHLIADYEIKNDLHVCFKSWFPAWAILAINKAGGAWPLDPSPHTQRHQIIVDQT 7993
QY 108 -A-A- -A- -A- 111
Db 7994 KARLALVSPNSITCIDLVEHVVEVSPVTDIELSKTSSHRGPREISPSNAAYVLTSG 8053
QY 112 --A--A--AA-- 115
Db 8054 STGTPKGLVMEHRSVCTSLTAITKLRPSARTLOFAAHVFDCAVGEIISLFTGGCLF 8113
QY 116 --A--A-- 117
Db 8114 VPSHDRMNLPEFIRQNKINYMSTPSTFRTLSPMVDPVPSLELVLVGEAVTRDMDTW 8173
QY 118 --A--S--Q-- 120
Db 8174 GKVLINGWGAETCVVSTTHEWOSIDESPLTIGRSGGFCWIVDPEDPQIAPVGAUGE 8233
QY 121 --Q--GS--A--K--N-- 126
Db 8234 IVIOGPTILREYLADPVRTSGSTVYNLPNAPNRDKYNNRKYSGDLGLYNADGTTIQFA 8293
QY 127 -- 126
Db 8294 SRKDTIRKIRGLRVELSEVHHVKTALSRPSCGCCRCIRSTRWCQVWPHTFALAMKQGL 8353
QY 127 --G--E-N--TA-- 131
Db 8354 LVLAEGFDECDPFTEIEMDNGMERLTAVISQINISLPRYMWPTFFIPCOYMPSTSRKLD 8413
QY 132 N--G--E--EN-G-- 137
Db 8414 NYLRRRTAALSQEQLVTSYLGKKRSPGTVMERILQAIWRSILCIPSENIGLDDSLG 8473
QY 138 --A-H-- 139
Db 8474 LGGDSITAIHLVMTSREGGISLTAQIDFDDPRLFAVARAKKAGVEGGSLEIPFPFSLLS 8533

QY 140 --T--IA--N-- 143
Db 8534 KRSQELVITDGVRAQCDLSSETIMDAYPCTALQEGLMALSVKQPGSYIAKYVYRIFSNVD 8593
QY 144 --N--HTD--M--M-- 149
Db 8594 IRRFKAAWQOTIALCDNLRITRIVLVDGRATQIVLBEPIAWTVSH-DLNTAVSEMRSVKMT 8652
QY 150 --E--V-DG-- 153
Db 8653 YGSPLSQFACIQNNNGDLFFLWSIIHAIIPDGWTVPIMLMTLHAFYRGIVTPPASPVARFV 8712
QY 154 --D-V--EI--PP--N--KA-V 163
Db 8713 KFAMSVDNEDTVKYWKQDLNNAKATFPFAASSQNSVSULEKSITCSQTAGSSITKATI 8772
QY 164 --VL--R-- 166
Db 8773 LRAAWALVLSRYSESDDVTFGATISGRQAPVDGITNMAGPAVATVPVRVICIDKKOTVAAF 8832
QY 167 --G--H-E--S--E--V-- 172
Db 8833 LDSVQSQALDMISYEQYGLQNISSKISSHAMEACDFGSLIIIOPIEOLAYRGDDALLVAAE 8892
QY 173 --F--I-C-- 175
Db 8893 SHITNETSALQNVFSYPLVIOCHLYDNTINMLIYDSQALSSESRMEALSHHPDNVQELA 8952
QY 176 --A-W--N-P-VS--D-L--LA-SG--S-- 188
Db 8953 SKONLLLSISLAGRWDFEQALKLNSDSDPAVERCFHEMDELTAIVRGDSLAIKSGDKSF 9012
QY 189 --GD--ST-- 192
Db 9013 TYREMAETNRIAHLVTDYGVKVGDIHVCPEKSAWFIATLAINKAGAAMSTLDPSPH 9072
QY 193 --A--R--I--W--N-- 197
Db 9073 TERYQKIINQTSQIALSSPVNSRCASLLPNVIELTSELDSKLAQNVWSTSRPAVNT 9132
QY 198 --L--S--E--N--ST-S-- 204
Db 9133 PRDAAYVLTSGTGVKPGVIEHGHSLCTSOISLKALDFNEBRVLQFSSSYFDILFE 9192
QY 205 --GST--Q--LV--LR-H-- 213
Db 9193 IGSTFLTGACLFVPSWDEQNNELVEYIRKHQLTFFMVLTTPLARTIRPEDVPVSDMLVVAG 9252
QY 214 --C--I-R--E-- 217
Db 9253 EAPTRDILDWFGKRLANGWGPTCECVIACLHQWTSVDESPKVIQKPIGGSCWIVDPED 9312
QY 218 --G-G--Q--D--VP--S-- 224
Db 9313 ATCMAPLGTGVEIIOGRNLLREYLSDPVKATATITGLPQWPKRDSIHWDRFYLTGDL 9372
QY 225 --N--KD-- 227
Db 9373 GFINEAGNLEYCTRKTQVKIRQORLELGEIHHIQANLESVROQAVDVIKSDAGSTLVA 9432
QY 228 --VT--S-LD-- 232
Db 9433 FVSFSDATQCTVTRVIANGPFTSLDGDFOATISRLGILSTLMPRYMVPSAFIPCAK 9492
QY 233 --W--N--N-- 234
Db 9493 GEFQHTGGRYKULRELAASQEDMAAYSLANEKVAQPMESRIQKIWAQVLNISIDS 9552
QY 235 --S-E-G--T--LIL-- 242
Db 9553 IGRDSDFLQGGSVSVIRLSVARDANIELTVGDIFFDPRLLAAKAKATEDWRDGLMT 9612

Qy 243 -----GS-----Y-----D-----G----- 247
Db 9613 QIBPFVLLDGSVKNVILSONIREQYSLPADKELEDAYPCTKLOEGLMALAVKQPGSYIAK 9672
Qy 248 FA-RI-----W-----T-----KD-----G----- 256
Db 9673 FLIRISPNVDVSHLRASWEETVRLPNLRTRIFTPNNMSVOAVFKDDVDVSWOSTSGTSLRS 9732
Qy 257 -----N--LA--S-----TL-----GOH-----K-----G----- 267
Db 9733 YMCSTENFQWYAGSPLARYALIEDGG-HTYPILSMHHAHVDGWMRVFMAVPHSIFHNRT 9791
Qy 268 -----P-----I-----P--A-----L-----K-- 273
Db 9792 HTVVEPYVRFVKYTLAIDSDAAADYMRSQFEGARQSIFPPTLSSVHEKKNSTLTLEKVI 9851
Qy 274 -----W-----N----- 274
Db 9852 DLPSTKSSITTATMLRAAWAIIILARYCDSDDITFGATISGRQAPVPGGLADMTGPAIATV 9911
Qy 275 -----N-----N-----K----- 276
Db 9912 PVRVWINRQIISDYLOAQIORQANDMIPFEQFGLQNIKSLSDAKACDFSSLLVTPIQ 9971
Qy 277 -----K-----G-----NFI----- 281
Db 9972 SLSYVDENADAIFVQAEVEKEIGTVQNYFSYPLVIQGHVHENFINLVIYDSNVLSNQ 10031
Qy 282 ---LS---A-----G-----VD---K--- 288
Db 10032 ITALSHQFESVMKQASQSDMKLGSVMSASDWDLEHSMRQNRDIPDVIDSCIHELIIQYQA 10091
Qy 289 -T-----T-----T-----I-- 291
Db 10092 VTQPDAPAIVSWDRDFTYKQLEASNRHLAHLVNKNVXKPDLLIPVFFEKANVYFAITA 10151
Qy 292 I-----W-----D----- 294
Db 10152 INKAGAAWVPLDPSHPVLRURQILSQTGTTTALSSANAVLCTLVKRVVNAELDKL 10211
Qy 295 -----AH-----TG-----E-----AK----- 301
Db 10212 LATESAHGPVVDVSRNAAYVLFTSGSTGIPKGLIMEHGSVCTSQVAIAKRLGLNSKVR 10271
Qy 302 -Q-----Q----- 303
Db 10272 ILQFAAFVPLDSIGVPLISGACICVPSEHIRKNSIANFINRQGITWTYLTTPSFVRTI 10331
Qy 304 -----F-----P-----F--H-----S--- 308
Db 10332 KASEVPNVKLLLLAGEAVPRDIFATWFGKRLINGWPAETCCFSTLHEMAICGRESYPC 10391
Qy 309 -----AP--AL-----DY-----DW--- 316
Db 10392 RRPVGFQWIVDPENPHRLAPTALGALGEVIIQGPITILREYLSVDVDRTEAAVIKSLPDWAPF 10451
Qy 317 --QS-----N----- 319
Db 10452 REQSSGRFNKSGDLGVNPDGTIEFSSRKDTQVKIRGLRVELGEVEHAQVALDGVHQI 10511
Qy 320 -----N-T-----F-----A--S----- 324
Db 10512 AVDVFKDNGTQPRGVFSFDESQIHEAHPSPFOAIDKQLARLTVAIGKLIALPRY 10571
Qy 325 -----C-----ST--D-----M-----C----- 330
Db 10572 MIPTLPICKMPSITSTKLDNRNLRRTLLTLQSELAMFSYRCNRKRTPTPTMESLIQRI 10631
Qy 331 -----IH----- 332
Db 10632 WSELHILPLDSIGRDSFLGLGDSITAIHLGSISSRQGISITAKDVFDPRLDWDIASKA 10691
Qy 333 -----V-----C-KL-G--- 337

Db 10692 RELDTAQQSPDISPFSFLIGETRELAAGSSVKQKLLTSTOTIEDAYPCTSKI PGSYI 10751
Qy 338 -----Q-D--R-----PI----- 342
Db 10752 AKYVYRLPHQVDIGRPFKSSWEHTVALHPILRTRIIMIKDTCVOLVVKDSTWENLDSGLD 10811
Qy 343 -----K-----T-----F--Q--- 346
Db 10812 DEAVNATHSYMTYGSPLSRYSIHESKSGDKYFLWTAHHSIHGWSVPVFNLTLYQAYKG 10871
Qy 347 -----G-----H----- 348
Db 10872 LEIGKPAVSGFKYKTWEHEQKAAGDYWRQLONAKRASPPKTGLAAKSIWTKRKRMS 10931
Qy 349 T-----NEV--NA--IK--W-----D-----P-----TG--- 361
Db 10932 TSLKFSASSNEIATKASVIRAAMAVILARYCESDDVCFGATISGRQASVPGLEMAPAV 10991
Qy 362 -----N-----LL--A-----S-----C----- 367
Db 10992 ATPVVRVOLDNDOEISKFLQNIQSOAHMVPEYQYGLQSIKLNAGDARDACDFTSLLVQ 11051
Qy 368 -----SD----- 370
Db 11052 PIQORQFVSQDDALLIPVETELEDILQSYNYNPLVQGHYNDYADLVLYDSTVVSEP 11111
Qy 371 -MT----- 373
Db 11112 QMTALCHFNFNIVQQLAAEDGKGLDISIASSFDLELAORSNGDGPQIIDDCIHIIER 11171
Qy 374 --K-----I--W-----SM----- 378
Db 11172 QAKQPNRPAIDAWDGKFTYSELDRTANRLAHLVHDYAVKVGDIHVHCFEKKWFFVAI 11231
Qy 379 ---KO---D-----N-----CV----- 384
Db 11232 LAVNKAGAAWAPPDPAHPORLKAVASQTGAKALASTANTRLCEQVWDCVVEVSTLKD 11291
Qy 385 -----H----- 386
Db 11292 NLWTVYDNIKGPDIINVTMDAAAYILFTSGSTGVPKGIVMQHGALCTNOAALSGWLGFDH 11351
Qy 387 ---LQ---Q-----H-----N----- 391
Db 11352 TVRMLQFSFVFDVSVEIVQALMNGACVCPSEHMRNLNSLDSFVRDFNVWTWYLTSPFT 11411
Qy 392 -----K-----E-----I-----Y-TI-KW----- 399
Db 11412 RTLKPKDPFSLKLLLAGPPTQDVLDTWFGLPNTRFINAWGPAETCVYNTLYEWQSNTE 11471
Qy 400 SP-----TG-----P-----G----- 405
Db 11472 SPLKGRAYGAYVWVDVENPQRLAFTGCLGEIIVQGPPLKEYLADPEKTAATVTELP 11531
Qy 406 -----T-N----- 407
Db 11532 EWAPRQSTWNRVYRTGDLGFDHGMHLFPASRKDTQVKIRGLRVELGEVEHRIQOGLD 11591
Qy 408 -----N-----P--N-----A-----NL--MLA-- 416
Db 11592 GVRQAVDVFKTEKGANLVAYLCFTSDTKTPQNTDFEGKDVFSADTDELQINLGRMLTE 11651
Qy 417 --SAS-----F-----D-----S----- 422
Db 11652 LNS-SLPTVYIPTMFIPCAVMPFITSSKLDVRKLRRLTAELSQQOLEAYSLVNEKOAPE 11710
Qy 423 T---VR---LW----- 427
Db 11711 TEMEIRLQKLWAEVLDPESIGRHNFMIRIGDSTIAAIRLVSMARDAGISLTVNDIFDD 11770
Qy 428 ----- 427

Db 11771 ARLISVATKAIDNEDSHLMAPTEPPSLLTSGIEDPALPSELSGQGFEDAYPCSKLQEG 11830
QY 428 -----DVP-----R----- 431
Db 11831 LMAIAIKQPGSYIAKYIKLSEHVDVDFKFAWESTVEIASALRTRIOTGGCSIQVIN 11890
QY 432 G-I-----C-----I-----H----- 436
Db 11891 GDSWDEADGSLQASQTLGMYGTRLCRYTLIRRDNDTYFLNMHHAVIDGLSTQN 11950
QY 437 ---TL-----T---K---H----- 441
Db 11951 ILGTLFNIYSGVDVLPUPPYNRFQYVLMQLEEDTANYWNQHLNQAQRTTFPPNDVARDK 12010
QY 442 -----Q---E-P-----V---Y---S---VAF---S----- 451
Db 12011 PAATQMLQSSIELPHGLDNSTVTIATVIRAAMAILVARYCSDSDVTGTTISGRQAPIPE 12070
QY 452 -----P-----D---G---R-----Y----- 456
Db 12071 IMGWGTIATVPVRVLINRQQLISDFLEGVQQAEMIAFEQYGLQNIAKLGDDARDAC 12130
QY 457 -----L---AS-G-----SF- 462
Db 12131 DFSSLLVQPIQHLAGTKDSDILVDASIGEDNAVEAIQNVFSYPLVIOAHLYDDHISFI 12190
QY 463 -----S---Y-----D-----K--- 464
Db 12191 LIYDSSIMLPQMVALSYLQVMMQLTNTPTATLETVSVSSYDVERALAINAEIPEVI 12250
QY 465 -C-----VH-----I----- 468
Db 12251 DTCHEMFENRVHLNPLAPAAWADGNYTYAELNSAANKLAHYLIKSYSKLANDLVHVCF 12310
QY 469 ---W---N---T-----QT-----G---ALV-- 477
Db 12311 DKSAWYIVSIIAINKAGATWPLDPHPPEQRLRSIVNQTKTLALASPGNITLCSALVDN 12370
QY 478 -----H----- 478
Db 12371 VVEVSSALDMLPAQDGLSPKVSVSRTAAIYLFSTSGTGTGKVVVQIHRSLCTNNAI 12430
QY 479 -----S-Y-----R-G----- 482
Db 12431 AKRVYHLDVRILOFSAYVDFPSILEIIMSLLQACVCPSEHIRMNGIYDFIRDMNINW 12490
QY 483 -----T-----GG-----I-- 486
Db 12491 LYLTPSLRTINPIDVPNVVELFAGGEAIPRDVFWGRVFRINGWGTETTIVGSIHE 12550
QY 487 FE-----V---CW-----N-----A 493
Db 12551 FESVDESPSTIGHVPGFCWIVDPNPNOLLAPTGLTGLGEIVIQGFTLLHLYLNDPNKTOEA 12610
QY 494 -----A-----GD-----K-----V-G----- 499
Db 12611 ILYDLPWAARPDENWGRFYKTDGLGFYNANGKIEPSSRKDTQVKIRGLRVELGEIEYQ 12670
QY 500 --AS-----A---SD-GS-V---C-----V-----LD----- 511
Db 12671 VQASVEIRQIADVKTNGSNLVAYLCFNFDEIRQLHNDVNGPFPSPLDNLQETLAGA 12730
QY 512 -----L-RK 514
Db 12731 IGKLSVTLPRYMIPTFFYIPCSYMPISITSGKLDK 12764

RESULT 15

Q6XA09

ID Q6XA09 PRELIMINARY; PRT: 7191 AA.

AC Q6XA09;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Nonribosomal peptide synthase.
GN Name:NRPS1;
OS Alternaria brassicae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.
OX NCBI_TaxID=29911;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14727058;
RA Guillemette T., Sellam A., Simoneau P.;
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria
RT brassicae and flanking genomic sequences.";
RL Curr. Genet. 45:214-224(2004).
CC -|- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL; AY246697; AAP78735.1; -;
DR InterPro; IPR010071; AA_adenyl_dom.
DR InterPro; IPR009081; ACP_like.
DR InterPro; IPR000873; AMP_bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR006163; Pp_bind.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; Condensation; 9.
DR Pfam; PF00550; PP-binding; 6.
DR PRINTS; PR00154; AMPBINDING.
DR TIGRFAMS; TIGR01733; AA-adenyl-dom; 4.
DR PROSITE; PS00075; ACP_DOMAIN; 7.
DR PROSITE; PS00455; AMP_BINDING; 2.
DR PROSITE; PS00697; DNA_LICASE_A1; UNKNOWN 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN 3.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN 3.
SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.2%; Score 2605.9; DB 2; Length 7191;

Best Local Similarity 8.6%; Pred. No. 3.5e-41;

Matches 452; Conservative 48; Mismatches 10; Indels 4751; Gaps 366;

QY 1 M-----S-----IS-----S---D-----E----- 7
Db 1521 MAFRLAGSRAPAGHAETVLARLWASVLKLDAGAVGAEDSPFRFGDSISAMKLVTAAR 1580
QY 8 -V-N---F---L---V---V---Y---R--- 14
Db 1581 KDGVLNVASVFAQPKLLAATAVMLPSDDAAKPEADTLPMLLPASSRQAIVALAAE 1640
QY 15 -----Y---LQE----- 18
Db 1641 CDVFPDCIEDMYPCSKLQGLVMTNKGDPGVVQPIYRLPADMDISRFRKAAKVVVAEE 1700
QY 19 -----S---GF-----S---H---S-----A-ET----- 27
Db 1701 ASLRTRIVYSEEHGLQVVVREDIKWQSLPDIQHINETRRRLPAKNGAPLATFTLVGENT 1760
QY 28 ---F-----G-----IK---S-H-----IS---QS-- 37
Db 1761 DSMFFVTAHAYVDCWSWTALFRKVENYRSGVHSPATVPYRSRVKYISSLDQOSDA 1820
QY 38 -----NI-----NGAL---V---P-----P---AA---L-- 49
Db 1821 FWLTQLDNVTAAQFPQLPGSPDHRVEANGQLLHTVLLTRNPGLEVTVFPMIRAAWGILLAT 1880
QY 50 -----I-SII-----OK----- 55
Db 1881 YSGSDVIVGNETNSGREASVPGIESIIGFTITPAVRLNRLNRLTVDHYLKETOKOSSLS 1940
QY 56 -----GLO-----Y-VE-----AE-----VS-I----- 66
Db 1941 LPYQFAGLQHIKSLSPETAICDFQSLGLEAGDDFDAESSLWNWVSANTIGTDFSYAF 2000
QY 67 -----N-----E-----D-----GT----- 71

Db 2001 VFCKVNAVQVEALLFDDRWQWLAQRWVOQDFILTLQNGTDNLRSLDLDLDP 2060
QY 72 -----L-----F-----D----- 74
Db 2061 DRKTISSWNSKVPPIIPRCHSVIAEDQTAIRLPTACAIADWDGFMYSYRELDERSALAH 2120
QY 75 -----G-RP-----I----- 78
Db 2121 QILRLGVRPKQFVPLCFDKSGMTVIAIIVLKAAPVPLDPEAPVLRLEIVSDIDADL 2180
QY 79 -----E-----S-----L-SL----- 83
Db 2181 LLCAPQYRELCSIPCSMTVVDRQATETVAGRLPSFVSHSDPAYAFVTSSTGPKPGA 2240
QY 84 -----I-----DA-V-M-----PD----- 90
Db 2241 VVHHTHWTSSTAFAGWKISTASRVLOFASYTFDACLIEVFSTLMQGGTVCVDPDGSRT 2300
QY 91 -----VV-----Q----- 93
Db 2301 NDLVGVINRNVNWAALTSPSVRMIVPSEIPOLETFLVGEAMSQDQDVTWADKVNNG 2360
QY 94 -----T-----R----- 95
Db 2361 YGTECAAVATSNIMTPHMRPNNLGRAVTARGWIVSRNNHHTLAPVGAIGELLEGAVG 2420
QY 96 -----Q-OA-----Y-R-D----- 101
Db 2421 AGYLNPEKTAQVFCQARWCVGLMGDDISAPVRIYKTDGLVKYEDGTWLYLGRKDLQT 2480
QY 102 -----KL-----A-Q-----Q----- 106
Db 2481 KYRGORLELSEVEHKLDDHVMQASALAVPTTGPCAKRLVAIVSLQKHGDHDDTDKQLR 2540
QY 107 -----Q-A-----A-----A-----A 113
Db 2541 LLPQENASLNIAIIRDCGLCERLPAYMIPSLWIAVERFPLMPGCKMRRCAIQWLEQMDQA 2600
QY 114 -----A-----A-----A-----AA----- 118
Db 2601 TYRLISAMGTDDAEVNGSPIEKLRALIAFAKVLNLSVADVRLNQSFHLGGSDIAAMQVS 2660
QY 119 SQ-OQ----- 122
Db 2661 SQCRAGFPISVDIIRSKSISAMASAVDSQSSQVTEAKDYDLFPDLSPIQKVFEDAV 2720
QY 123 -----SA-----KN-G-----E----- 128
Db 2721 GDKHNYFNQTELPRLSRNITEBELRSALTATLAKTHPMLRARFSEKNEAGAWKQRIEKDVSS 2780
QY 129 -----NTAN-----C-----E----- 134
Db 2781 SFRLRHHVQAGNDANLRPIIDHSQATLDVAKGFTFAIELFDVDDTFSSQIALVAHHLII 2840
QY 135 -----EN-----G-A-----H-----T-----IA----- 142
Db 2841 DVVSGILLEDLQGLLQGLQAPQSPMPYHMLHEQSLQATQESARRVFPVGDIAQDLDY 2900
QY 143 -----N-N-----H-T-D-WM-----E-V-D----- 152
Db 2901 WAMEGRPNVNGDVVEEDLHLSTRDTMLLLQAQDALATETIDILIAALFESFRKVPFDRST 2960
QY 153 -----G-----D-----V-E-----I-----P- 158
Db 2961 IAHNEGHGRETFNNRQDLSRTVGWFTVTPHLPVPLNEATDMISTIRWDRFRNRTPD 3020
QY 159 -----P-----NK----- 161
Db 3021 KGRPYFAYRNLTGQTRFASHWPAEAVFNVYGRLOSQDNKQGLFTALSDVDSREVGEDV 3080
QY 162 -----V-----AV----- 164
Db 3081 PRLALFDITAAVSGAIIKLSFGWNRNMRQKEIRAWVGKCRQTLVDAVEELLQARQERSV 3140

QY 165 -----L-----R-----G-----H-E----- 169
Db 3141 GNFKYLPLLYNGISRLSAILPAGINLVNVEDIYPASPMQOGLLHTQSRHPELYTYHTVSQ 3200
QY 170 -----S-----E-----VFI----- 174
Db 3201 VQSADGNPIDPRRLAEAWQVWVHRHQALRTIFIDSLAKDGSKQDQIVLKEKPGRVQILADC 3260
QY 175 -----C-A-----W-----N----- 178
Db 3261 DDSQVANLLRHOSIDCREALPPHRMSICKTKTRVWFKLELSHVINDGTSVSNLLADLA 3320
QY 179 -----P-----V-----S-DL-L-A-SG-----S 188
Db 3321 RAYARKLTRADAGPLYSDIYGWMLSRSSDADLAYWKAHLGSGIEPCLFPVLDNDGIPSSPES 3380
QY 189 G-D-----STARI-----W-----N 197
Db 3381 GSVDELGSTRVQDFCKQNGVTLSNVQLTALTLHYVVTGTPDVSFGLIASGRDIPVTN 3440
QY 198 LSE-----N-----S-T-----SGST-----OL----- 209
Db 3441 INEAVGCFVNMVMSRLSFSDETTIAQLLEALQGTSTEALSHQCSLADIQHALLQPSLFN 3500
QY 210 -----V-----L-R----- 212
Db 3501 TAFTPORRSLSDPEDTALIYEDMEADAGEYIVTVNADVTQDSITVDGYSKDRILPSQ 3560
QY 213 -----H-----C-----I----- 215
Db 3561 AQNMAETFKILDSIVVCSASELTIGKLDILTESSIHQIMWNPQPPPIRRCLHDVHD 3620
QY 216 -----R-----EG-G-----QD----- 221
Db 3621 QALTRPRTTKAVEGMDGTFTYQDFDKITNQLAVHLOSIGVTTTETFPILFEKSSYAIVSM 3680
QY 222 ----- 221
Db 3681 IAIMKAGGAYVPLDPKHPQTRRLRELIEDVGASVVLCSRGYHTTASEVAKTAVIVDQBSIR 3740
QY 222 -----VP-S-----N-----K----- 226
Db 3741 KLGVPISSPRTCATPDNAAYCLFTSGTTGPKGTIIIPHOAFCTSAATFRRMINATSR 3800
QY 227 -----D-----VTS-L-----D-----W----- 233
Db 3801 TFOFASYTFDASCIETLSALTATVGVPTEDDRMNAAGAIRKLRVNMSSLLTPSVLGTI 3860
QY 234 ----- 235
Db 3861 EBERVPLKTLVSGGEALSGPILKKNWSNCTCFINAYGPTCSVVAATAYKSTLDHKLIVS 3920
QY 236 E-GTLATGS-----Y-----D-----G-FAR-----I----- 251
Db 3921 EFGT-IGTSGGRLWIVHPRNHDKLMVGVSGELVIEGTARGYLNDVKTAKAFINDP 3979
QY 252 -W-----T-----K-----DG-NL-----AST-----L-QO-----H----- 265
Db 3980 AWAKTIFSSNNTFEARMYKTGDLVRYNTDGSVNYIGRKDTQIKLNGQRIELGEIBFHV 4039
QY 266 ----- 265
Db 4040 KNPPERVQSAVELVAPNSRSSAKALAVFAVQDQIDGEOVSQVQASTDLPAADLLPL 4099
QY 266 -----K-----G-----P-IF-----ALW----- 274
Db 4100 SDELDMCKNTENGLAGSLPSYMIPIAFIPVTKLPWTSAGKLDNRNLSRVQNLSTW 4159
QY 275 -----NKK-----G----- 278
Db 4160 MYRLTSIANKKKPITEAEKKIHKAVCVLSLPPSSVIGDSDSVRLGDSSTSSMRLVAMAH 4219

QY 279 -----N-----F-----I----- 281
Db 4220 TEQWELSFIDIFKNPKSLDLAKTGAQISKSQAEKWQPPDLLPASLTRSDVSEVVQOC 4279
QY 282 -----LS--AG-----V-----DKT----- 289
Db 4280 QVSKEDLQDAYPTSSLQDALLTSLIKQAGAYVAQHLALPKSLDMTKFKAAWESAIOEID 4339
QY 290 --T-II-----W-----DA-----H-----T-- 297
Db 4340 ILTRIIQMPGIFMOAVLRENDPVDWREAKSLKSAEDDASKIPPHLGHLAAYTLVTTPS 4399
QY 298 GE-----A-----K-----Q----- 302
Db 4400 GERFYFWTLHHALYDGSIIYMLQVQOIIYSGKGVSTTPQTSYARFVEYLSSTVSUSDVY 4459
QY 303 -----QPP-----F-HS--A-----PA----- 311
Db 4460 WRERLTGVNAYQFPRPSHATSPAPPNGQMFQHSKMTAHRKNTDVTTPANATRAAWALILAA 4519
QY 312 -----L-----DV----- 314
Db 4520 YTGSDDDVFGETLAGRDVAMTGDVCGPTLTTPVSRVKIDRGATVSDLLNTIATNIDR 4579
QY 315 -----S-----N-----D-W-----Q----- 317
Db 4580 IPYQHGLSAIKALGEMIAACDFQNLVLIQTENEELADSMWSVHONEEQGNFTYPLVI 4639
QY 318 -----S-----N-----T-----F-----AS----- 324
Db 4640 ECKWGLSKTEVLAHFDANVISLHWQRLIYQFETVLIQLOSATHVRHIAVLSDQDKQVLR 4699
QY 320 --N-----T-----F-----AS----- 324
Db 4700 KWAYEPRLDDTVPSLFFKKAASQPTTAVTAFDGEFSYSGELASALQAELVKFGAG 4759
QY 325 -----C-----S-TD----- 328
Db 4760 PECLIPVDKSRWAVAILAILISGAGVYPLSPTDPASRHLHIVETCKASIVLCSPKYT 4819
QY 329 -----M-----C-----I-H----- 332
Db 4820 HRFVEMGVHVSSETAIRQLPTSSISLSQRAKNNICVIFTSGTGLPKGVVIEHXS 4879
QY 333 -----VCKLG-----Q-----D-----R-----P----- 341
Db 4880 VSSAAICE-GLHITPTSRVFQFCFLFVSVGETLTVLLRGATICVPSDEQRTNLAA 4938
QY 342 IKT-----F-----QG-----HTN----- 350
Db 4939 V-TDLNANWAFLPVSNAVSTLEGPKSVPTLETLVVGGEAMTSDVVDKWATGVNLH-NGYGP 4996
QY 351 -E--V-----N-----A-----I-----K-----W-----D-----P-----T----- 360
Db 4997 TEGTVFAIGNDHVSAQRDPNSIGHPLKSGRAWLTNSDNPHELAPIGATABELCLEGPILLAR 5056
QY 361 G--N-----L-----L-----A-----S----- 366
Db 5057 GYLNDPKRTSEAFIAPFLKNFSSNESRIYRTGDLVQYAADGSIQYMGKRNQIKLAG 5116
QY 367 -----C-----SD----- 369
Db 5117 QRIELDEIQAVVHADNNHVQVQLPKVGPCCKLTVVVSFPGTAAASAGSDHRRILSDT 5176
QY 370 -----D-----M-----TL-----K----- 374
Db 5177 ESLSQINRARDRLADLPVSVMVPIWIAVPRIPTLASAKLDKQVGLWLEGMEALYQRI 5236
QY 375 -----TW-----S-MK----- 379
Db 5237 MGAELPEDMEGPGAAALTVLIRGIWAKVLRNRPVEDVPKSWLSLGGDSISAMKLLAKCRS 5296
QY 380 -----Q-D----- 381

Db 5297 EGINLNLQILRAKSLSHLAADVKSVVILDHEQNDRPFALSPIORFYVEAGSIENSTH 5356
QY 382 -----N-----C-----V----- 384
Db 5357 FNOGSTLRILRYVQPAWQALNSIVECHSMLRARFSPKONNGOWOOLVMSKVSQSYAFTA 5416
QY 385 HD-----L--Q-----O-----H----- 390
Db 5417 HDVSTASAGAISSQKSLDRTGPVFAVDLNLKGHQLFMAAHLHVIDVVSWGILLG 5476
QY 391 -----N----- 391
Db 5477 DLEDLLGSGPVTLPRSLPFQWCKMQTSSASEITQOLTVKNOPLVVEPANPAFWMGDVDR 5536
QY 392 -----KE-----IY-T-----I-K----- 398
Db 5537 PNYGDSERDEFVIDKETSAMAFDNHHVYKTDLVDILLAAILHSFBSRVFINRKAPTLFNE 5596
QY 399 -----W-----S-----P-----T-----G-P-- 404
Db 5597 SHGREVWDGNSNLDLSRTGVWFTTLYPVTVPIDDEDEVIHTLRQVDTTRKVASNGRPYF 5656
QY 405 -----G-----TN-----N-----P----- 409
Db 5657 AHRYLTEDGKERFANHPMEVLPNLYGRQBSGHSLSLPTQVEGDDDETSVDGVKTSR 5716
QY 410 -----LAS-----A--SPD-----L-----M-- 414
Db 5717 MALFEISASVTEGOIQLSPMYNRYSKNKGIRRWIAECORTLEAIELAKTRDPOPTWA 5776
QY 415 -----LAS-----A--SPD-----ST-- 423
Db 5777 DPPLPLESYRLDRVLKTLPHAGVSPFOVEDMPCSPIDGMILSOIKSPESYSSTT 5836
QY 424 --VR-----L--W-----D-----V-DR-----GI--C 434
Db 5837 FEVRKRGFVDASKVVDGKQVVARHPALRTIPIDSVCKGVFDQVWKSQSDGIVTYKC 5896
QY 435 -----I-HT-----L-----T-----K-----H----- 441
Db 5897 ADAELATLLESIRHSLNGKKKPVLPQAAVQVOTSSGKIFVKIVNVHAVIDGSGSLGVIGQ 5956
QY 442 --QE-----PVYS-----VA-----F-----S-P-- 452
Db 5957 DLOEAYEGRLESDGPYISAYIKYLRALPAEDAIAYWAKALRGVSPCYFFTPTRDPKPRQ 6016
QY 453 -----D-----G-----R-----YLAS----- 459
Db 6017 LQSLDMRFTFRDELHDLAESSNVTIANILLAAWALVRSYTNSSDVCYGLTSGRNVPID 6076
QY 460 -----GSF-----D-----K-----C----- 465
Db 6077 RIENAYGAFINMLVSRIELKSATSLLEIVENQSDFGVSGMPHQHCSLAOPQHDLGLSGKS 6136
QY 466 --V-----H----- 467
Db 6137 LFNATVSIQRRSPBELTSDSGIBFEQLDCHDPSEFAITVNIDATEDDEGVEFTYWSDA 6196
QY 468 ----- 467
Db 6197 VTDGEAKNVSTLMAKILVQVLSNPQKOTIAELDIWVKRPAQPAVSRNLNPKRPSILRSS 6256
QY 468 -----I-----W-----N-----N-- 470
Db 6257 SSISRSSTGPPRPTITFPDLAPAAPLPAETPDWSSLIRSIIVSEMVPOIIVEQIVAKNLS 6316
QY 471 -----TQ-TCAL-----VHSVR----- 481
Db 6317 TEPTSATIDMTQMTQMLTRKTSISQGRGPSIDTAGSPRAASIHRRASIASNAENRI 6376
QY 482 -----G-----T-----GG--I 486

Db 6377 QTAADMVATLGVLATEASSKVAPDFVEKLLNLWAELELLEMVETIEQDDSFQLOGDSII 6436
QY 487 -----F-----EV-C----- 490
Db 6437 AMELVGNAAREEGLSMTVADVFNKPTFADMARVVVRVAGEVIDEVMRSRAGGSVAGRSAGGQS 6496
QY 491 -----W-----NA----- 493
Db 6497 RSRHLDRAPSIWSEFQDIMSDPNADTKSIAPSEMPAEPGLNSRDSTMFORWQGLTTNQAR 6556
QY 494 -----A-----KV-----G----- 499
Db 6557 PAASREVSKSLATHTIQEGVSSVNRSVSMLGDPNVESVISKQVFKGGISDVFPVTFD 6616
QY 500 ---A---S-----A----- 502
Db 6617 QSLAITGLMESKWLNYFYLDGDPDLRLKLAQAAAYRMVHAFDILRTVFPYGDRLQV 6676
QY 503 -----SD-----G----- 505
Db 6677 VLRKLQPEFYHQTDIDTFTKDLRKQKORENGPKLGEAFIQFWAKQKQOTGRYRIFMRL 6736
QY 506 S-----VC---VJ-----D---L-----R---K 514
Db 6737 SHAQYDGVCMKILNALQDGYNGLPVSSAPSPGPNFVRETAK 6777

Search completed: January 3, 2005, 15:25:01
Job time : 135.333 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:53:45 ; Search time 95.3333 Seconds
(without alignments)
1934.128 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 3669

Sequence: 1 MGISDEVNPLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1: Genesecp1980s:*

2: Genesecp1990s:*

3: Genesecp2000s:*

4: Genesecp2001s:*

5: Genesecp2002s:*

6: Genesecp2003as:*

7: Genesecp2003bs:*

8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3669	100.0	514	8 ADJ76263	Adj76263 Marker ge
2	3622	98.7	514	8 ADJ75513	Adj75513 Marker ge
3	3613	98.5	514	4 AAB95225	Aab95225 Human pro
4	3612	98.4	514	6 ABO07190	Ab007190 Human p53
5	3350.2	91.3	577	7 ADD14051	Add14051 Human src
6	3350.2	91.3	577	8 ADQ18019	Adq18019 Human sof
7	3260.6	88.9	542	5 ABP41760	Abp41760 Human ova
8	3250.7	88.6	459	4 ADM19860	Admi9860 Protein e
9	2916.9	79.5	700	4 ABB60376	Abb60376 Drosophil
10	2791.1	76.1	584	4 ABG21351	Abg21351 Novel hum
11	2716	74.0	395	5 ABP51424	Abps1424 Human MDD
12	2631.9	71.7	5635	5 ABP60991	Abp60991 Novel hum
13	2627.8	71.6	5636	7 ADJ70089	Adj70089 Human hea
14	2627.8	71.6	5636	7 ADJ83137	Adj83137 Human hem
15	2627.8	71.6	5636	8 ADK60205	Adk60205 Angiogene
16	2627.8	71.6	5636	8 ADK60506	Adk60506 Angiogene
17	2627.8	71.6	5636	8 ADP73129	Adp73129 Angiogene
18	2617.2	71.3	7064	7 ADGP70546	Adgp70546 Aspergill
19	2612.9	71.2	4599	8 ADI27168	Adi27168 Mouse LRP
20	2612.9	71.2	4599	8 ADI27169	Adi27169 Mouse LRP
21	2609	71.1	31267	6 ABG74786	Abg74786 Human RGS
22	2608.1	71.1	4796	4 ABB58665	Abb58665 Drosophil
23	2594.4	70.7	4599	6 ABP56837	Abp56837 Human LRP
24	2593.4	70.7	4636	4 AAE11937	Aae11937 Human CGI
25	2592.2	70.7	26926	4 AAU05396	Aau05396 Human tit

ALIGNMENTS

RESULT 1

ADJ76263
ID ADJ76263 standard; protein; 514 AA.

XX AC ADJ76263;

DT 20-MAY-2004 (first entry)

DE Marker gene related amino acid sequence SEQ ID NO:1515.

XX KW bronchial asthma; chronic obstructive pulmonary disease;

KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;

XX KW gene therapy; marker.

XX OS Mus musculus.

XX PN EP1394274-A2.

XX PD 03-MAR-2004.

XX PF 04-AUG-2003; 2003EP-00254857.

XX PR 06-AUG-2002; 2002JP-0029312.

XX PA 20-MAR-2003; 2003JP-00077212.

XX XX (GENO-) GENOX RES INC.

XX PI Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuwara K;

XX DR WPI; 2004-193155/19.

XX PT Testing for bronchial asthma or chronic obstructive pulmonary disease by comparing the expression level of a marker gene in a biological sample from a subject with the expression level of the gene in a sample from a healthy subject.

XX PS Claim 16; SEQ ID NO 1515; 241pp; English.

XX CC The present invention describes a method of testing for bronchial asthma or chronic obstructive pulmonary disease. The method comprises determining the expression level of a marker gene in a biological sample from a subject, comparing the expression level determined with the expression level of the marker gene in a biological sample from a healthy subject, and judging whether the subject has bronchial asthma or chronic obstructive pulmonary disease. The marker gene comprises: (a) a group of genes (S1) whose expression levels increase when respiratory epithelial cells are stimulated with interleukin-13; or (b) a group of genes (S2) whose expression levels decrease when respiratory epithelial cells are

Adq17316 Human sof
Adn96832 Bugula br
Adn96829 Bugula br
Abb11353 Human LDL
Aau74797 Mouse alp
Adn11590 Human CD9
Adn11588 Human CD9
Adn11587 Human CD9
Adn11589 Human CD9
Aar47861 Alpha 2-M
Aar60517 Human alp
Aam9091 Human pro
Aau81019 Human alp
Abp56839 Human LRP
Abu89744 Protein d
Add14025 Human src
Add14025 Human src
Adi27167 Human LRP
Adi15636 Human lip
Adn11584 Human CD9
Abb58144 Drosophil

26 2592.2 70.7 26926 8 ADQ17316
27 2590.2 70.6 5509 8 ADN96832
28 2590 70.6 6940 8 ADN96829
29 2588.1 70.5 4563 4 ABB11353
30 2587.7 70.5 4545 5 AAU74797
31 2586.1 70.5 4419 8 ADN11590
32 2586.1 70.5 4419 8 ADN11588
33 2586.1 70.5 4464 8 ADN11587
34 2586.1 70.5 4464 8 ADN11589
35 2586.1 70.5 4544 2 AAR47861
36 2586.1 70.5 4544 2 AAR60517
37 2586.1 70.5 4544 4 AAM9091
38 2586.1 70.5 4544 5 AAU81019
39 2586.1 70.5 4544 6 ABP56839
40 2586.1 70.5 4544 6 ABU89744
41 2586.1 70.5 4544 7 ADD14025
42 2586.1 70.5 4544 8 ADI27167
43 2586.1 70.5 4544 8 ADI15636
44 2586.1 70.5 4544 8 ADN11584
45 2585.5 70.5 7107 4 ABB58144

CC stimulated with interleukin-13. Also described: (1) a reagent (I) for
 CC testing for bronchial asthma or chronic obstructive pulmonary disease;
 CC (2) a kit for screening for a candidate compound for a therapeutic agent
 CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
 CC an animal model for bronchial asthma or chronic obstructive pulmonary
 CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
 CC method for producing an animal model for bronchial asthma or chronic
 CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial
 CC asthma or chronic obstructive pulmonary disease, comprising the compound,
 CC a marker gene or an antisense nucleic acid corresponding to a portion of
 CC the marker gene, a ribozyme, a polynucleotide that suppresses the
 CC expression of the gene through an RNAi effect or an antibody recognising
 CC a protein encoded by a marker gene; and (7) a DNA chip for testing for
 CC bronchial asthma or a chronic obstructive pulmonary disease, on which a
 CC probe has been immobilised to assay a marker gene. (1) has respiratory
 CC and antiasthmatic activities, and can be used in gene therapy. The method
 CC is useful for testing for or screening for a therapeutic agent for
 CC bronchial asthma or chronic obstructive pulmonary disease. The present
 CC sequence is used in the exemplification of the present invention.
 XX
 SQ Sequence 514 AA;

Query Match 100.0%; Score 3669; DB 8; Length 514;
 Best Local Similarity 100.0%; Pred. No. 5.9e-56;
 Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLOESGFSHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60
 DB 1 MSISDEVNLFVRYLOESGFSHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAATNQ 120
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAATNQ 120
 QY 121 QGSAKNGENTANGEGAHTIANNHTDMVEVDGVEIPSNKAVVLRGHESEVFICAWNPV 180
 DB 121 QGSAKNGENTANGEGAHTIANNHTDMVEVDGVEIPSNKAVVLRGHESEVFICAWNPV 180
 QY 181 SLLVSGSGSTARIWNLSENSTSGTQLVLRHCIRREGGQDVPSNKDVTSLDWNSEGLL 240
 DB 181 SLLVSGSGSTARIWNLSENSTSGTQLVLRHCIRREGGQDVPSNKDVTSLDWNSEGLL 240
 QY 241 ATGSDYGFARITWKGDLASTLQHGKPIIFALKWNKXGNFILSAGVDKTTIINDAHTGEA 300
 DB 241 ATGSDYGFARITWKGDLASTLQHGKPIIFALKWNKXGNFILSAGVDKTTIINDAHTGEA 300
 QY 301 KQPFPHSAPALVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 DB 301 KQPFPHSAPALVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSKQDNCVHDLOAHNKEIYTIKWSPTGPTNPNANMLLASASF 420
 DB 361 GNLLASCSDDMTLKIWSKQDNCVHDLOAHNKEIYTIKWSPTGPTNPNANMLLASASF 420
 QY 421 DSTVRLMDVRGICHTLTKHQBVSVAFSPDGRYLAGSPDKCVHIIWNTQTGALVHSY 480
 DB 421 DSTVRLMDVRGICHTLTKHQBVSVAFSPDGRYLAGSPDKCVHIIWNTQTGALVHSY 480
 QY 481 RTGGIPEVCWNAAGDKVGASDGSVCVLDLRK 514
 DB 481 RTGGIPEVCWNAAGDKVGASDGSVCVLDLRK 514

RESULT 2
 ADJ75513
 ID ADJ75513 standard; protein; 514 AA.
 XX
 AC ADJ75513;

XX 20-MAY-2004 (first entry)
 XX
 DE Marker gene related amino acid sequence SEQ ID NO:765.
 XX

KW bronchial asthma; chronic obstructive pulmonary disease;
 KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;
 KW gene therapy; marker.
 XX
 OS Homo sapiens.
 XX
 PN EPI394274-A2.
 XX
 PD 03-MAR-2004.
 XX
 PF 04-AUG-2003; 2003EP-00254857.
 XX
 PR 06-AUG-2002; 2002JP-00229312.
 PR
 PR 20-MAR-2003; 2003JP-00077212.
 XX
 XX (GENO-) GENOX RES INC.
 XX
 PI Ontani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;
 DR WPI; 2004-193155/19.
 XX
 PT Testing for bronchial asthma or chronic obstructive pulmonary disease by
 PT comparing the expression level of a marker gene in a biological sample
 PT from a subject with the expression level of the gene in a sample from a
 PT healthy subject.
 XX
 XX Example 11; SEQ ID NO 765; 241pp; English.

XX The present invention describes a method of testing for bronchial asthma
 CC or chronic obstructive pulmonary disease. The method comprises
 CC determining the expression level of a marker gene in a biological sample
 CC from a subject, comparing the expression level determined with the
 CC expression level of the marker gene in a biological sample from a healthy
 CC subject, and judging whether the subject has bronchial asthma or chronic
 CC obstructive pulmonary disease. The marker gene comprises: (a) a group of
 CC genes (S1) whose expression levels increase when respiratory epithelial
 CC cells are stimulated with interleukin-13; or (b) a group of genes (S2)
 CC whose expression levels decrease when respiratory epithelial cells are
 CC stimulated with interleukin-13. Also described: (1) a reagent (I) for
 CC testing for bronchial asthma or chronic obstructive pulmonary disease;
 CC (2) a kit for screening for a candidate compound for a therapeutic agent
 CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
 CC an animal model for bronchial asthma or chronic obstructive pulmonary
 CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
 CC method for producing an animal model for bronchial asthma or chronic
 CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial
 CC asthma or chronic obstructive pulmonary disease, comprising the compound,
 CC a marker gene or an antisense nucleic acid corresponding to a portion of
 CC the marker gene, a ribozyme, a polynucleotide that suppresses the
 CC expression of the gene through an RNAi effect or an antibody recognising
 CC a protein encoded by a marker gene; and (7) a DNA chip for testing for
 CC bronchial asthma or a chronic obstructive pulmonary disease, on which a
 CC probe has been immobilised to assay a marker gene. (1) has respiratory
 CC and antiasthmatic activities, and can be used in gene therapy. The method
 CC is useful for testing for or screening for a therapeutic agent for
 CC bronchial asthma or chronic obstructive pulmonary disease. The present
 CC sequence is used in the exemplification of the present invention.

XX Sequence 514 AA;

Query Match 98.7%; Score 3622; DB 8; Length 514;
 Best Local Similarity 98.8%; Pred. No. 4.6e-55;
 Matches 508; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLOESGFSHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60
 DB 1 MSISDEVNLFVRYLOESGFSHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAATNQ 120
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAATNQ 120
 QY 121 QGSAKNGENTANGEGAHTIANNHTDMVEVDGVEIPSNKAVVLRGHESEVFICAWNPV 180

Db 121 QGSAKNGENTANGEANGHTIANNHTDMMEVDGVEIPNKAIVLRGHESEVFICAWNPV 180
QY 181 SLLVSGSGDSTARINWLNSENSTGPTQLVLRHCIREGGQDVPSNKKDVTSLDWNSEGTLL 240
Db 181 SLLVSGSGDSTARINWLNSENSTGPTQLVLRHCIREGGQDVPSNKKDVTSLDWNSEGTLL 240
QY 241 ATGSYDGFARITWKDGNLSTLQHKGPFPALKWKKGNFPLSAGVDKTTIWDHTGEA 300
Db 241 ATGSYDGFARITWKDGNLSTLQHKGPFPALKWKKGNFPLSAGVDKTTIWDHTGEA 300
QY 301 KOQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
Db 301 KOQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANLMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANLMLASAF 420
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSFCKVHIWNTQTGALVHSY 480
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSFCKVHIWNTQTGALVHSY 480
QY 481 RGTGGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514
Db 481 RGTGGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514

RESULT 3
AAB95225
ID AAB95225 standard; protein; 514 AA.
XX AC AAB95225;
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:17352.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.
The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX SQ Sequence 514 AA;

Query Match 98.5%; Score 3613; DB 4; Length 514;
Best Local Similarity 98.6%; Pred. No. 6.8e-55;
Matches 507; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MSISSDEVNFLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIKQLQYV 60
Db 1 MSISSDEVNFLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIKQLQYV 60
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTQQAYRDKLAQHQAAAAAATNQ 120
Db 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTQQAYRDKLAQHQAAAAAASQ 120
QY 121 QGSAKNGENTANGEANGHTIANNHTDMMEVDGVEIPNKAIVLRGHESEVFICAWNPV 180
Db 121 QGSAKNGENTANGEANGHTIANNHTDMMEVDGVEIPNKAIVLRGHESEVFICAWNPV 180
QY 181 SLLVSGSGDSTARINWLNSENSTGPTQLVLRHCIREGGQDVPSNKKDVTSLDWNSEGTLL 240
Db 181 SLLVSGSGDSTARINWLNSENSTGPTQLVLRHCIREGGQDVPSNKKDVTSLDWNSEGTLL 240
QY 241 ATGSYDGFARITWKDGNLSTLQHKGPFPALKWKKGNFPLSAGVDKTTIWDHTGEA 300
Db 241 ATGSYDGFARITWKDGNLSTLQHKGPFPALKWKKGNFPLSAGVDKTTIWDHTGEA 300
QY 301 KOQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
Db 301 KOQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANLMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANLMLASAF 420
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSFCKVHIWNTQTGALVHSY 480
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSFCKVHIWNTQTGALVHSY 480
QY 481 RGTGGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514
Db 481 RGTGGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514

RESULT 4
AAB007190
ID ABO07190 standard; protein; 514 AA.
XX AC ABO07190;
XX DT 13-AUG-2003 (first entry)
XX DE Human p53 modifying protein, SEQ ID 150.
XX KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic; antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;

lung cancer; ovarian cancer; angiogenesis; cell cycle;
apoptotic disorder; cell proliferation disorder.
Homo sapiens.
WO200299122-A1.
12-DEC-2002.
03-JUN-2002; 2002WO-US017382.
05-JUN-2001; 2001US-0296076P.
10-OCT-2001; 2001US-0328605P.
15-FEB-2002; 2002US-0357253P.
(EXEL-) EXELIXIS INC.
Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
WPI; 2003-156859/15.
N-PSDB; ACD13365.
Identifying modulators of the p53 pathway for use in treating apoptotic
or cell proliferation disorders, comprises screening for agents that
modulate activity of a human ortholog of genes that modify the p53
pathway in Drosophila.
Example 2; Page 458-459; 678pp; English.
The invention relates to identifying (M1) a candidate p53 pathway
modulating agent, by contacting an assay system comprising a purified HM
polypeptide (human orthologue of genes that modify the p53 pathway in
Drosophila) or nucleic acid with a test agent under conditions, where but
for the presence of the test agent, the system provides a reference
activity, and detecting a test agent-biased activity of the assay system.
Also included are modulating (M2) a p53 pathway of a cell (comprising
contacting a cell defective in p53 function with a candidate modulator
that specifically binds to a HM polypeptide comprising an HM amino acid
sequence, where p53 function is restored), modulating (M3) a p53 pathway
in a mammalian cell (comprising contacting the cell with an agent that
specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
a disease in a patient (comprising: (a) obtaining a biological sample
from the patient; (b) contacting the sample with a probe for HM
expression; (c) comparing the results with a control; and (d) determining
whether the comparison indicates a likelihood disease). (M1) is useful
for identifying modulators of the p53 pathway. A probe for HM expression
in a patient, where the cancer has greater than 25 % expression level.
Modulators identified by (M1) are useful in a variety of diagnostic and
therapeutic applications, where disease or disorder prognosis is related
to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
proliferation disorders (e.g. cancer). Another two new methods (M2 and
M3) are useful for modulating the p53 pathway of a cell, thus restoring
the p53 function of the cell, so that the cell undergoes normal
proliferation or progression through the cell cycle. (M2) and (M3) are
also useful for treating defects in the p53 pathway such as angiogenic,
apoptotic or cell proliferation disorders. The present sequence
represents a human p53 pathway modifying protein
Sequence 514 AA;
Query Match 98.4%; Score 3612; DB 6; Length 514;
Best Local Similarity 98.6%; Pred. No. 7.1e-55;
Matches 507; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MSISDSEVNLVRYLQESGFSHSAFTFGIESHSQSNGALVPPAALISIIKGLQYV 60
Db 1 MSISDSEVNLVRYLQESGFSHSAFTFGIESHSQSNGALVPPAALISIIKGLQYV 60
QY 61 EAEVSINEDCTLFDGRPIESLSLIDAVMPDVOTROQAYRDLKLAQOHAHAHAHAATNQ 120
Db 61 EAEVSINEDCTLFDGRPIESLSLIDAVMPDVOTROQAYRDLKLAQOHAHAHAHAASQ 120

121 QGSAKNGENTANGEANGAHTIANNHTDMMEDVDGVEIPSNKAVVLRGHESEVFCANPV 180
121 QGSAKNGENTANGEANGAHTIANNHTDMMEDVDGVEIPPNKAVVLRGHESEVFCANPV 180
181 SLLVSGSGDSTARINWLSNSTSGPTOLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
181 SLLVSGSGDSTARINWLSNSTSGPTOLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
241 ATGSYDGFARITKDGKLNASTLGQHKGPFPALKWKKGNFIFLSAGVDKTTIWDHTGEA 300
241 ATGSYDGFARITKDGKLNASTLGQHKGPFPALKWKKGNFIFLSAGVDKTTIWDHTGEA 300
301 KQFPFHSAAPALVDWQSNNTFASCSTDCIHVKLGQDRPIKTFQGHNTNEVNAIKWDPT 360
301 KQFPFHSAAPALVDWQSNNTFASCSTDCIHVKLGQDRPIKTFQGHNTNEVNAIKWDPT 360
361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKIYTIKWSPTGPTNNPNANLMLASAF 420
361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKIYTIKWSPTGPTNNPNANLMLASAF 420
421 DSTVRLWDVDRGICHTLTKHQEPVSVAFSPDGRYLASGSPDKCVHIWNTOTGALVHSY 480
421 DSTVRLWDVDRGICHTLTKHQEPVSVAFSPDGRYLASGSPDKCVHIWNTOTGALVHSY 480
481 RGTGGIFEVCAAGDKVGASASDGSVCVLDLRK 514
481 RGTGGIFEVCAAGDKVGASASDGSVCVLDLRK 514

RESULT 5
ADD14051
ID ADD14051 standard; protein; 577 AA.
XX AC ADD14051;
XX DT 01-JAN-2004 (first entry)
XX DE Human src biomarker polypeptide SEQ ID NO:240.
XX KW predictor set; protein tyrosine kinase activity modulator;
XX KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
XX KW gene therapy; drug sensitivity; genetic profile; cancer; human.
XX OS Homo sapiens.
XX PN WO2003062395-A2.
XX PD 31-JUL-2003.
XX PF 17-JAN-2003; 2003WO-US001981.
XX PR 18-JAN-2002; 2002US-0350061P.
XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX PI Huang F, Fairchild CR, Lee FY, Shaw P;
XX DR WPI; 2003-636735/60.
XX DR N-PSDB; ADD14646.
XX PT New polynucleotides and polypeptides for predicting the activity of
XX PT compounds that interact with protein tyrosine kinases and/or protein
XX PT tyrosine kinase pathways.
XX PS Claim 10; SEQ ID NO 240; 139pp; English.
XX CC The present invention describes a predictor set comprising a plurality of
XX CC polynucleotides or polypeptides whose expression pattern is predictive of
XX CC the response of cells to treatment with a compound that modulates protein
XX CC tyrosine kinase activity or members of the protein tyrosine kinase
XX CC pathway. Also described: (1) predicting whether a compound is capable of
XX CC modulating the activity of cells, comprising obtaining a sample of cells,
XX CC determining whether the cells express a plurality of markers, and


```

PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Rosen CA, Barash SC, Ruben SM;
XX
PI WPI; 2001-476159/51.
DR N-PSDB; ADM19381.
XX
PT Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Claim 11; SEQ ID NO 667; 809pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule encoding a
CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to a protein of the
CC invention.
XX
SQ Sequence 459 AA;
Query Match 88.6%; Score 3250.7; DB 4; Length 459;
Best Local Similarity 91.9%; Pred. No. 3.1e-48;
Matches 452; Conservative 3; Mismatches 4; Indels 33; Gaps 3;
XX
QY 23 HSAFTFGIESHSQSNINGALVPPAALISIIQKGLQVVEAVSINEDGTLFDRPIESLS 82
Db 1 H-A-----SD-----VEAEVSINEDGTLFDRPIESLS 27
QY 83 LIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 142
Db 28 LIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 87
QY 143 NNHTDMNEVDGVEIPNKAVALRGHSEVFIQWNPVSDLLVSGSDSTARIWNLSNS 202
Db 88 NNHTDMNEVDGVEIPNKAVALRGHSEVFIQWNPVSDLLVSGSDSTARIWNLSNS 147
QY 203 TSGPTQLVLRHCIREGGQVPSNKNVTSLDWNSGTLATGSDGFGARITWKDGNLSTL 262
Db 148 TSGSTQLVLRHCIREGGQVPSNKNVTSLDWNSGTLATGSDGFGARITWKDGNLSTL 207
QY 263 GQHKGPFIKLNKNGKFNILSAGVDKTTIWDATGEAKQFPFHSAPALDWDVQSNNTF 322
Db 208 GQHKGPFIKLNKNGKFNILSAGVDKTTIWDATGEAKQFPFHSAPALDWDVQSNNTF 267
QY 323 ASCSTDMCIHVCKIGDPRPKTFQHTNEVNAIKWPTGNLLASCDDMTLKIWSMKQDN 382
Db 268 ASCSTDMCIHVCKIGDPRPKTFQHTNEVNAIKWPTGNLLASCDDMTLKIWSMKQDN 327
QY 383 CVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAFDSTVRLWDVDRGICHTLTTHQ 442
Db 328 CVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAFDSTVRLWDVDRGICHTLTTHQ 387

```

```

QY 443 EPVYSVAFSPDGRYLASGSPDKCVHIWNTOTGALVHSYRGTTGGIFVCWNAAGDKVGASA 502
Db 388 EPVYSVAFSPDGRYLASGSPDKCVHIWNTOTGALVHSYRGTTGGIFVCWNAAGDKVGASA 447
QY 503 SDGSVCVLDLRK 514
Db 448 SDGSVCVLDLRK 459

```

```

RESULT 9
ABB60376
ID ABB60376 standard; protein; 700 AA.
XX
AC ABB60376;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 7920.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL04479.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 7920; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABLI01840-ABLI6175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 700 AA;

```

```

Query Match 79.5%; Score 2916.9; DB 4; Length 700;
Best Local Similarity 57.3%; Pred. No. 5.4e-41;
Matches 411; Conservative 51; Mismatches 34; Indels 221; Gaps 61;
QY 1 MSISDEVNLFVRYLOESGSHSAFTFGIESHSIQSNINGALVPPAALISIIQKGLQV 60
Db 1 MSFSDEVNLFVRYLOESGSHSAFTFGIESHSIQSNINGALVPPAALISIIQKGLQV 60
QY 61 EAEVSINDEGTLFD-GRPIESLSLIDAMPDV-----VQTRQ-----AYPD---KLA-- 104
Db 61 EVESVSGEDG---EVAPRIEGLSLIDAMPVEVKPLIVKT-EFGKPGAV-DSSAP-AGG 114
QY 105 -Q-----QH-----A-A-----A-----A-----AA---A-----A 115

```

Db 115 NQNNNAKPEIKBPGTGAGSAGNKIAGSTTGTSTPTDQASGEVDSGNAANAGTVA 174
QY 116 ---A---T---N-----Q-----QSAK--N-G-EN--T-- 130
Db 175 GNNAGAGNQASTGSGNSTPTAGGDLAAPGASQKQNSNEAGSSGNAAGNATSTDD 234
QY 131 -A-----NG-----EE-----N-----G-AH--TI- 141
Db 235 AASSTSTNGSSSTSSVQEQPTSGLTGAGTGTSTNPDAAASGASATGSKASGAVTIR 294
QY 142 --A--NN-----H-----T-DMVEY--DGD--VEIP-SNKAVLVR 166
Db 295 VQAQGNVQSGSSNAQSSAPSGTSSSTSGGAGTPAAL-VPMIDENIEIPES-KARVLR 352
QY 167 GHESEVFIKAWNPV--DLLVSGSGDSTARIWNLE-NSTSGPTQLVLRHCIRGGQDVPS 224
Db 353 GHESEVFIKAWNP-SRDLASGSGDSTARIWDSANTNS--NQLVLRHCIRGGQAEVPS 409
QY 225 NKDVTSLDMNSEGTLLATGSDYGFARIW-TKQGNLSTLQHGKPIFALKWKK-GNFIL 282
Db 410 NKDVTSLDMNCDGSLATGSDYGVARIWKT-DGRLASTLQHGKPIFALKW-KCNYIL 467
QY 283 SAGVDKTTIWDHATGE-AKQPPFHSAPALVDVDSNNFTASCST-D-MCHVKLG--Q 338
Db 468 SAGVDKTTIWDASTGQCT-QPFAFHSAPALVDVQTNQAFASCSTQD-IHVCRLGVNE 525
QY 339 DRPIKTFQHTNEVNAIKWDPKTNLLASCSDMTLKIWSMKQD-NCVHDLQAHNKIYTI 397
Db 526 --PIKTFKHTNEVNAIKWCPQQLASCSDMTLKIWSMNRDCC-HDLQAHKSIYTI 582
QY 398 KWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTQKQBPVYSVAFSPQRYL 457
Db 583 KWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTQKQBPVYSVAFSPQRYL 642
QY 458 ASGSFQKCVHIWNTQALVHSYRGTTGPIEVCWNAAGDKVGSASDGSVCVLDLRK 514
Db 643 ASGSFQKCVHIWNTQALVHSYRGTTGPIEVCWNSKGTQVGSASDGSVCVLDLRK 699

RESULT 10
ABG21351
ID ABG21351 standard; protein; 584 AA.
XX AC ABG21351;
XX DE 18-FEB-2002 (first entry)
XX DT Novel human diagnostic protein #21342.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX XX WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX XX (HYSE-) HYSEQ INC.
XX PA Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS85538.
XX DR New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensic, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 20; SEQ ID NO 51710; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. The invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 584 AA;
Query Match 76.1%; Score 2791.1; DB 4; Length 584;
Best Local Similarity 63.1%; Pred. No. 5.6e-39;
Matches 383; Conservative 44; Mismatches 41; Indels 139; Gaps 52;
QY 2 SI-----S--S-----D-----EV---N---FLV-YRY-LQE-SCFSSHATFGIESH 33
Db 22 ATGAVGFRQAATLQAGPDGNGVCVHAGNSRIF-VPLPVGL-EVKGFSHATFGIESH 79
QY 34 ISOSNINGALVPPAALISITIQGLQYVEAEVSNEDGTLFDGRPIESLSLIDAVMPDVQ 93
Db 80 ISOSNINGTLVPPSALISILQGLQYVEAEISINKDGTVDSPRIESLSLIVAVDPVQ 139
QY 94 TRQAYRDKLAQQA--A--A--AAT-----NQGSAXKNGENTANGENGATHTA 142
Db 140 MRQAFGEKLTQQAASAAATEASAKAATMTPAALISQNPVKREATVNGEANGAEI- 198
QY 143 NNHTDM-MEVDGDVEIPSNKAVLRGHESEVFCAMNPVSDLLVSGSGDSTARINLSEN 201
Db 199 NNHS-KPMEIDGDVEIPPNKATVLRGHESEVFCAMNPVSDLLVSGSGDSTARINLSEN 257
QY 202 STSGPTQLVLRHCIRREGQDVPSNKDVTSLDWNSEGTLLATGSDYGFARIWTKDGNLST 261
Db 258 SNGSGTQLVLRHCIRREGHDVPSNKDVTSLDWNSDGTLLAMGSDYDGFARIWTE--N-A-- 312
QY 262 LGQHKGPFIAP--KW-NKXGN--FILS-AGVDKTTIWDHAT--GEAKQKQFPFHSAPAL 312
Db 313 -----P--ALDWDQW---NMTFA-SCS-TD---HVCRLGCD-----H--P-- 344
QY 313 DVD-WQSNNTFASCSTDMCIH-VC-----KL-----G-QDRPIKTF-----QG- 347
Db 345 -VKTFQ-GHTF--C-T--CIESICFWGGLRKLTTMTTEGKRLRP-KTFCSDGGALLPPAGR 396
QY 348 --H--T-----NEVNAIKWDPKTNLLASCSDMTLKIWSMKQDNCVHDLQA 389
Db 397 RPHLLTGPDIFPKRLSALFQNEVNAIKWDPKTNLLASCSDMTLKIWSMKQDNCVHDLQA 456
QY 390 HNKEIYTIKWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTQKQBPVYSVA 449
Db 457 HSKEIYTIKWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTQKQBPVYSVA 516
QY 450 PSPDGRYLAGSFDKCVHIWNTQALVHSYRGTTGPIEVCWNAAGDKVGSASDGSVCV 509
Db 517 PSPDGRYLAGSFDKCVHIWNTQALVHSYRGTTGPIEVCWNSKGTQVGSASDGSVCV 576

DT 10-SEP-2002 (first entry)
 XX Novel human protein. SEQ ID 78.
 XX
 KW Human; cytostatic; vulnary; antiarteriosclerotic; antiparkinsonian;
 KW neurotropic; neuroprotective; immunosuppressive; haemostatic;
 KW antiinflammatory; cardiant; antiulcer; virucide; antithyroid;
 KW cerebroprotective; anorectic; metabolic; vaccine; cancer; infection;
 KW wound healing disorders; atherosclerosis; Parkinson's disease;
 KW Alzheimer's disease; autoimmune disorder; haematopoietic disorder;
 KW inflammation; neoplastic disease; nervous system disorder;
 KW cardiovascular disorders; pancreatitis; respiratory disorder;
 KW hyperproliferation; systemic autoimmune disease; hyper-immunity;
 KW developmental abnormality; gastrointestinal ulceration; neuropathy;
 KW haematological disease; metabolic disease; sperm dysfunction;
 KW thyroid disorder; hypothyroidism; brain damage; colitis;
 KW cone photo- transduction deficiency; neurological disease; stroke;
 KW angiogenesis; ovulation disorder; spinal cord; thyroid gland; heart;
 KW trachea; thymus; lymph node; muscular system; obesity; anorexia;
 KW growth abnormality; precocious puberty.
 XX Homo sapiens.
 XX WO200250105-A1.
 XX
 XX 27-JUN-2002.
 XX
 XX 17-DEC-2001; 2001WO-US049232.
 XX
 XX 19-DEC-2000; 2000US-0256710P.
 XX 20-DEC-2000; 2000US-0257048P.
 XX 09-JAN-2001; 2001US-0260482P.
 XX 30-JAN-2001; 2001US-0264922P.
 XX 06-FEB-2001; 2001US-0266797P.
 XX 19-MAR-2001; 2001US-0276988P.
 XX 04-APR-2001; 2001US-0281535P.
 XX 08-MAY-2001; 2001US-0289622P.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 XX (GLAX) GLAXO GROUP LTD.
 XX
 XX Agartwal P, Birkeland M, Cogswell JP, Kabnick KF, Lai Y;
 XX Martensen SA, Rizvi SK, Smith RF, Strum JC, Xie Q;
 XX
 XX WPI: 2002-508784/54.
 XX N-PSDB; ABQ86156.
 XX
 XX Secreted proteins and polynucleotides useful as vaccines for preventing
 XX or treating various diseases e.g. cancer, wounds, atherosclerosis,
 XX Parkinson's disease, Alzheimer's disease, infection, autoimmune disorder.
 XX
 XX Claim 1(a); Page 285-297; 335pp; English.
 XX
 XX The invention relates to an isolated polypeptide with signal sequences
 XX which allow it to be secreted extracellularly or membrane associated. The
 XX activity of polypeptides of the invention may be described as,
 XX cytostatic, vulnary, antiarteriosclerotic, antiparkinsonian, neurotropic,
 XX neuroprotective, immunosuppressive, haemostatic, antiinflammatory,
 XX cardiant, antiulcer, virucide, antithyroid, cerebroprotective, anorectic,
 XX and metabolic. Polypeptides and polynucleotides of the invention are
 XX useful in the treatment, or as a vaccine in the prevention of, cancer,
 XX wound healing disorders, infection, atherosclerosis, Parkinson's disease,
 XX and Alzheimer's disease, autoimmune disorder, haematopoietic disorder,
 XX inflammation, neoplastic disease, nervous system related disorders and
 XX cardiovascular disorders, pancreatitis, respiratory disorder,
 XX hyperproliferation, systemic autoimmune disease, hyper-immunity,
 XX developmental abnormality, gastrointestinal ulceration, neuropathy,
 XX haematological diseases, metabolic diseases, sperm dysfunction, thyroid
 XX disorders e.g. hypothyroidism, brain damages, colitis, cone photo-
 XX transduction deficiency, neurological diseases, stroke, angiogenesis,
 XX ovulation disorders, diseases in the spinal cord, thyroid gland, heart,
 XX trachea, thymus, lymph node and muscular system, obesity, anorexia,

CC growth abnormalities, and alleviation of precocious puberty. The
 CC sequences given in records ABP60965-ABP61019 represent novel human
 CC proteins of the invention
 XX
 XX Sequence 5635 AA;
 XX
 XX Query Match 71.7%; Score 2631.9; DB 5; Length 5635;
 XX Best Local Similarity 9.6%; Pred. No. 3.3e-31;
 XX Matches 440; Conservative 57; Mismatches 15; Indels 4071; Gaps 372;
 XX
 QY 1 M-S-----I-----SS--D-----BV-----8
 DB 1 MISWEVHTVFLFALLYSLAQDASQSEIRAEIEPEGASTLAFVFDVTGSMYDDLQVVI 60
 QY 9 -----NF-LV-----12
 DB 61 EGASKILETSLKRPRKPLNFALVPHDPDPEIGPVITTPDKPKFQYELRELYVGGDCPE 120
 QY 13 -----Y-----R-----Y-----LQ-----17
 DB 121 MSIGAIIKIALEISLPGSFIYVFTDARS KDYRLTHEVLQIIQKQSQVFLVTGDCDRTH 180
 QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
 DB 181 IGYKVEEIASTSSGQVFLDKKQVNEVLKQVVEAAVQASKVHLLSTDHLEQAVNTWRIPF 240
 QY 27 -----T-----FG-----I-----E-----31
 DB 241 DPSLKEVTVSLGSPSPMIEIRNPLKLIKKGGLHLLHNSAKVVNVKEPEAGMWTVK 300
 QY 32 -----S--H-----I-----S--Q-----36
 DB 301 TSSSGRHSVRITGLSTIDFRAGSRKPTLDFKKTVSRRPQGIPTYVLLNTSGISTPARID 360
 QY 37 -----S-----S-----NI-----N-----G-----41
 DB 361 LLELLISGSLKTIPTVKYYPHRKPYGIWNISDFVPPNEAFFLKVTGYDKDDYLFQVSS 420
 QY 42 -----A-----L-----VP-----P-----A-----47
 DB 421 VSFSSIVDPAPKVTMPKTPGGYLPQGPICSVDSLLPFTLSFVRNGVTGLVDQYLKESA 480
 QY 48 -----A--L-----I-----S--IIQ-----KG-----56
 DB 481 SVNLDIAKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPPIQVNNVTVTPGERAVLT 540
 QY 57 -----L--Q-----Y-----V--B-----61
 DB 541 CLIIISAVDYNLTQWRNDRVRLAEPARIRTLANLSLELKVKNFNDAGEYHCVMSSEGGSS 600
 QY 62 -A-----EVS1-----NE-----68
 DB 601 AASVFLTVQEPKVTMPKQKQSFQSGSEVSMCSATGYPKPKIATVNDMFIVGHSHRYM 660
 QY 69 --DGLTF-----D-----G-----R-----P-----I-- 78
 DB 661 TSDGTLFIKNAAPKAGIYGCILASAGTDKQNSTLYIEAPKLMVQSELLVALGDITV 720
 QY 79 -B--S-----L--S--LI-----DA-----V-----87
 DB 721 MECKTSGIPPPQVQWFKGDLRLPSTFLIDPLLLGLLKIIQETQDLADAGDYTCVAINAEGR 780
 QY 88 -----M-----P-DV-----VQ-----T-----94
 DB 781 ATGKITLVGSPPPVFIQEPADVSMIEIGSNVTLPVYVQGPETIKWRRLDNMPISRPFS 840
 QY 95 -----R-----95
 DB 841 VSSIISQIRGALFILNWSKGTICEAENQFGIKQSETTIVTGLVAPLIGISPSVAN 900
 QY 96 -----QQ-----A-----Y-----R-D-KL-----AQ-----105
 DB 901 VIEGQQLTLPCLLAGNPPIERRWIKNSAMLLQNPNYITVRS DGSGLHIERVOLQDGGEYTC 960

QY 106 -----OH-----A----- 108
Db 961 VASVAGTNNKTTVVVHVLPITQHQOILSTIEGIPVTLPCASGNPKPSVWISKKGEL 1020
QY 109 -----A-A-A-----A-A-A----- 114
Db 1021 ISTSAKESAGADGSLYVSPGEBSEYVCTATNTAGYAKVKQLTVYVPRVFGDQRG 1080
QY 115 -----A-----A----- 116
Db 1081 LSQDPVBISVLAGEBVTLPCEVKSPLPPIITWAKETQLISFPSPRHFTLPSSGSMKITET 1140
QY 117 -----ATN-----Q-----Q-GSA-K-----N----- 126
Db 1141 RTSDSGMYLCVATNAGNVTQAVKLVNHVPPKIQRGPKHLKVQVQGVQVDPICNAQGTPLP 1200
QY 127 -----GE-N-----T-A-N-G----- 133
Db 1201 VITWSKGSWLVDEHHVSNPDGTLSDQATPSDAGIYTCVATNAGTDETEITLHVQE 1260
QY 134 -----E-N-----GA-----H----- 139
Db 1261 PPTVEDLEPPYNTTFOERVANORIEFFPCPAKGTPTIKWHLNGRELTCREPGISILEDG 1320
QY 140 T--IA-----N-----N--HT--DM----- 148
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKKEQVTVNSVLLNQLTNL 1380
QY 149 -----M-----E----- 150
Db 1381 PCEVEGTPSPIMVYKDNVQVTESTIQTVNGKILKILFRATPEDAGRYSCAKAINAGTS 1440
QY 151 -----VD-----G-----DV-B-----I-----P- 158
Db 1441 QKYNIDVLPPTTIIGNFNEVSVLNRDVALECOVKGTPTFDIHWFKDGKFLFLDPN 1500
QY 159 -----SN--K-A-----V-----VL 165
Db 1501 VELLDRGOVLHKNARENDKRYOCTVSNAGKQAKDKILTYNPPSIKGNVTTDISVL 1560
QY 166 -----RG-----H----- 168
Db 1561 INSLIKLECETRGLPMPAITWKDQGPIMSSQALYIDKQYLHPIRAQVSDSATYCHV 1620
QY 169 -----E-S-----E-----VFI-----C-A-----W----- 177
Db 1621 ANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECKAAGNPSPIILTKDGV 1680
QY 178 -----N-----P-V-SD- 182
Db 1681 PVKANDNIRIEAGGKLEINSAQIDRGQYICVATSVAGEKEIKEYVDVLVPPAIEGGDE 1740
QY 183 -----LL-----VSGS-----G-----D-----S-- 191
Db 1741 TSYFVIMVNNLLELCHVTGSPPTIMWLKDGOLIDERDGFILLNGRKLVIQAQVSN 1800
QY 192 -----TA-----R-----I-W 196
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSSGLSERVVVYKVPVALQCIANGIPNPSITW 1860
QY 197 -----NL-----S-----E-----NS--T-----S 204
Db 1861 LKDDQPNTAQGNLKIQSSGRVLQIAKTLLEDAGRYTCVATNAAGETQQHQLHVHEPPS 1920
QY 205 -----G-----PTOL-----V-----L-R----- 212
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVEVITWYKDNRLLSGTSMTFLNRQIIDIS 1980
QY 213 -----H-----C-----I-- 215
Db 1981 AQISDAGIYKCVAINSAGATELFYSLQVHVAPSIGSNMNVAVVNNPVRLSEARGIPA 2040

QY 216 -----R-----E-----G-----G-Q--D-- 221
Db 2041 PSLTWLKDGPVSPFSNGLQVLSGGRILALTSQISDGTGRYTCVAVNAAGEKQRDIDLRV 2100
QY 222 -VPSN-----KD-----VT--S-L- 231
Db 2101 YVPFNINGBEQNVSVLISQAVELLCSDAIPPTLTWLDKGHPLKPKGLSISSENKSVLK 2160
QY 232 -----D-----W-N--S-----EG--TLL--A 241
Db 2161 IEDAQVQDTRYTCATNVAGTEKNYVNVNWPPIGSGDELQTLQTVIEGNLISLCE 2220
QY 242 T-----GS--Y-D-----G-----F-----A-- 249
Db 2221 SGIPPNLIWKKKSPVLTDMSGRVILSGGROLQISIAEKSDAALYSCVASNVAGTAK 2280
QY 250 -----R-----I-----I-----WKDG----- 256
Db 2281 EYNLQVIRPTITNSGSHPTIIVTRKGSISLECEVOGIPPTVTWMDGHPLIKAGVE 2340
QY 257 -----NL-AS-----T-----L-----GOH----- 265
Db 2341 ILDEGHILQKNIHVSQDTRYVCVAVNAGMTDKKYDLSVHAPPSIIGNHRSNPENISVVE 2400
QY 266 K-----G--P-I--FALK--W----- 274
Db 2401 KNSVSLTCEASGIPLESITWF--KDGWPVSLNSVRILSGRMLRLMOTTMEDAGQYTCV 2458
QY 275 -----N-----K-K-----GN----- 279
Db 2459 VRNAGEERKIFGLSVLVPPIHVGENTLEDVKVKEQSVTLTCEVTGNPVPITWHKQOQ 2518
QY 280 -----F-I-----L-----S--AGV-- 286
Db 2519 PLQEDEAHHIISGGRFLQITNVQVPHGTGRYTCCLASSPAGHKRSFSLNVFSPITAGVGS 2578
QY 287 D--K-T-----T-I-----I-W----- 293
Db 2579 DGNPDEVILNPSITSLVCEAYSPPATITWPKOGTPLESNRNIRILPGGRTLQILNAQE 2638
QY 294 D-----A-----H-----TG-----BA 300
Db 2639 DNAGRYSCVATNEAGEMIKHYEVKVIYPIINKGDLWGPGLSPKEVKIKVNNLTILECA 2698
QY 301 -----K-----Q-----Q-----F 304
Db 2699 YAIPASLSWYKDGQPLKSDDHVNIANGHTLQIKEAQISDTGRYTCVASNIAGEDELDF 2758
QY 305 -----P-F-----H-- 307
Db 2759 DVNIQVPPSFQKLWEIGNWMLDTRNGEAKDVIIINNPIISLYCETNAAPPPTLTWYKDGHP 2818
QY 308 -SA-----P-----A-----LDV-----D- 315
Db 2819 TSSDKVLILPGGRVLIQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIKANGS 2878
QY 316 -----WQ-----SN----- 319
Db 2879 PEEVTVNKSALIECLSSGSPAPRNSWQDQGPLLEDHKKFLSNGRILQILNTQITDI 2938
QY 320 -----NT-----FAS--C-----S- 326
Db 2939 GRVYVCAENTAGSACKYFNLVNHPVPSVIGPKSENLTVVVNNFISLTCEVSGFPDLSW 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 LKNEQIKLNTNLIVPGGRTLQIIRAKVSDGGEYTCIAINQAGSKKFSILTVVPPSI 3058
QY 332 -H-----V--C-----K-----L--QO----- 338
Db 3059 KDHSSESLSVNVNREGTSVLECESNAVPPPVITWYKGRMITESTHVEILLADGQMLHIK 3118
QY 339 -----D-----R-----P-IK-----TF----- 345

CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON) mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, neurotropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cyostatic activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.

XX
 SQ Sequence 5636 AA;

Query Match 71.6%; Score 2627.8; DB 7; Length 5636;

Best Local Similarity 9.6%; Pred. No. 3.9e-31;

Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;

QY 1 M-S-----I-----SS--D-----EV-----8
 DB 1 MISWEVHTVFLFALLYSSLAQDASQSEIRAEFFPEGASTLAFVFDVTGMYDDLQVVI 60
 QY 9 -----NF-LV-----12
 DB 61 EGASKILETSLKRPKRLFNALVPHDPEIGPVITITDPKFKQYELRELYVOGGDCPE 120
 QY 13 -----Y-----R--Y-----LQ-----17
 DB 121 MSIGAIAKLEISLPGSFIVFTDARSKDYRLTHEVLQLIQKQSQVVFVLTGDCDRTH 180
 QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
 DB 181 IGYKVEEIASTSGGVFHLDDKQVNEVLKWBEAVQASKVHLLSTDLHLEQAVNTWRIPF 240
 QY 27 -----T-----T-----FG--I-----E-----31
 DB 241 DPSLKEVTVSLSGSPMIEIRNPLGLIKKFGHLLHNHNSAKVNVKEPEAGWTVK 300
 QY 32 -----S--H-----I-----S--Q-----36
 DB 301 TSSSGRHSVRITGLSTIDFRAGFSKPTLDFKKTVRPVGQIPTVYLLNTSGISTPARID 360
 QY 37 -----S-----S-----NI-----N-----G-----41
 DB 361 LLELSISGSLKTI PVKYYPHRKPGYIWNISDFVPPNEAFFLKVTDYKDYLFORVSS 420
 QY 42 -----A-----L-----VP-----P-----A 47
 DB 421 VSFSSIVDPAPKVTMPKTPGYLQPGQIFCSVDLSLLPFTLSFVRNGVTILGVQYLKESA 480
 QY 48 -----A-----L-----I-----S-----IIO-----KG-----56
 DB 481 SVSLDIKAVTLSDGFEVCIASVAGSAGRAQTFDVSPPPPVQVPPNVTVTGERAVLT 540
 QY 57 -----L--Q-----L-----Y--V--E-----61
 DB 541 CLIIASVDYNTQWRDRVRLAEPARITLANLSLELKSVPKNDAGEYHCWVSSEGGSS 600
 QY 62 -A-----EVSI-----NE-----68
 DB 601 AASVELTVQBPVKVMPKNQSTGGSEVSMCSATCYPKPKIAWTNDMFIVGSHRYRM 660
 QY 69 --DGTLP-----D-----G-----R-----P-----I--78
 DB 661 TSDGTLFIKNAKPDAGIYGLASAGTDKQNSTLRYIEAPKLMVQSELLVALGDITV 720
 QY 79 -E--S-----L--S--LI-----DA-----V-----87
 DB 721 MECKTSGIPPPQVKNFKGDLRLPSTFLIIDPLLLGLLKIQTODLDAGDYTCVAINAAGR 780

QY 88 -----M-----P-DV-----VQ-----T-----94
 DB 781 ATGKIILDVGPSPPVFIQEPADWSNEIGSNVTLPCYQGVPEPTIKWRLDNMFISRPFS 840
 QY 95 -----R-----95
 DB 841 VSSISQLRTGALFILNWDKGTIYCEARNQFKIQSETTVTVTGLVAPLIGISPSVAN 900
 QY 96 -----QQ-----A-----Y-----R-D-KL-----AQ-----105
 DB 901 VIEGQQLTLPCTLLAGNPPIPERWRKNSAMLLQNPYITVRSDGLSHIERVQLQDGGSEYTC 960
 QY 106 -----OH-----A-----108
 DB 961 VASNVAGTNNKTSVVVHVLPITIQHQQLLSTIEGIPVTLPCKASGNPKPSVIMSKGEL 1020
 QY 109 -----A--A--A-----A--A--A-----114
 DB 1021 ISTSSAKFSAGADGSLVWSPEGESGEYVCTATNTAGVAKRKVQLTVVVRPVFGDLRG 1080
 QY 115 -----A-----A-----116
 DB 1081 LSQDKPVEISVLAGEBVTLPCEVKSLLPPIITWAKETQLISPPSPRHTFLPSSGMKITET 1140
 QY 117 -----ATN-----Q-----Q-GSA--K-----N-----126
 DB 1141 RTSDSGMYLCVATNIAGNVTQAVKLVHVPPKIQRGPKHLKVQGVQGVQVDPICNAQGTPLP 1200
 QY 127 -----GE--N-----T--A--N--G-----133
 DB 1201 VITWSKGGSTMLVDGHHVSNPDGTLSDQATPSDAGIVTCVATNIAGTDEITLHVQE 1260
 QY 134 -----E-----E--N-----GA-----H-----139
 DB 1261 PPTVELEPPYNTTFOERVANQRIEPPCPAKGPKTKIWLHNGRELTGREPGISILEEG 1320
 QY 140 T--IA-----N-----N--HT-----DM-----148
 DB 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKQEVSNVSVLLNQLTNL 1380
 QY 149 -----M-----E-----150
 DB 1381 FCEVEGTPSPIIMWYKDNQVVTESSTIQTWNNGKILKLFATPEDAGRYSCKAINIAGTS 1440
 QY 151 -----VD-----G-----DV--E-----I-----P--158
 DB 1441 QKYFNIDVLVPPTIIGTNFPKEVSVVLRNDRDALEQCQVKGTPPPDIIHFKDGKPLFLGDPN 1500
 QY 159 -----SN--K-A-----V-----VL 165
 DB 1501 VELLDRGQVLHLKNARRNDKGRVQCTVSNAAGKQAKDIKLTIVIPPSIKGNNVTTDISVL 1560
 QY 166 -----RG-----H--168
 DB 1561 INSILKLECKTRGLPMPAITWYKDGQIPMSSSQALYIDRGQYLHI PRAQVSDSATYTCVH 1620
 QY 169 -----E-S-----E-----VFI-----C--A-----W-----177
 DB 1621 ANVAGTAESFHVVDVVPVPMIEGNLATPLNKQVVIASHLTLECNAAAGNPSPIITWLKDG 1680
 QY 178 -----N-----P-V--SD--182
 DB 1681 PVKANDNFRIEAGGKKLEIMSAQIDRGQYICVATSVAGEKEIKYEVDVLVPAIEGGDE 1740
 QY 183 -----LL-----VSGS-----G-----D-----S--191
 DB 1741 TSYFIVMNNLELDCHVTGSPPTIMWLKDGOLIDERDGFKILLNGRKLVIATAQVSN 1800
 QY 192 -----TA-----R-----I--W 196
 DB 1801 GLYRCMAANTAGDKKKEFEVTVHPPTIKSSGLSERVVVKYKVALQCIANGIPNFSITW 1860
 QY 197 -----NL-----S-----E-----NS--T-----S 204

QY 467 -----H-----I-W----- 469
Db 4079 PPVISPHEKVIYADVADKPTLSCADGLPPDITWHKDGRAIVESINQRVLSGSLQIAF 4138
QY 470 -----N-----T-----Q----- 472
Db 4139 VQPDAGHYTCMAANVAGSSSTKLTVHVPPRIIRSTEGHYTVNENSOAILPCVADGPT 4198
QY 473 -----T-----G-----A-----L-VH 478
Db 4199 PAINWKKDNVLANLLGKYTAEPYGELILENVVLEDSGFVTCVANNAAAGDTHVSLTVH 4258
QY 479 -----S-----Y-----R 481
Db 4259 VLPTFTLPGDVSINKGEQLRSLCKATGIPLPKLTWTFNNNIIPAHFDSVNGHSELVIER 4318
QY 482 -----GT-----G-----IF-----E-----V--C----- 490
Db 4319 VSKEDSGTYVCTAENSVGFKATGFFVYVKEPPVFKGDPNSNWIPLGGNAILNCEVKGDP 4378
QY 491 -----WN-----A-----AGD----- 496
Db 4379 TPTIQNRKGVDEISHRIQLGNSLAIYGTVNEADGDTVCATNEAGVVERSMSTLQ 4438
QY 497 -----K-V-----G-----A-----SAS----- 503
Db 4439 SPPIITLPEVTVINAGGKIILNQATGEPQPTITWSRQHSISDWRVNLNNSLYIA 4498
QY 504 -----D-----GSV-----C-V----- 509
Db 4499 DAQKEDISEFCVARNLMGSLVLRVPVIVQVHGFSQWSAWRACSVTCGKGIQKRSCLN 4558
QY 510 --L-----DL--R-----K 514
Db 4559 QPLPANGKPCQGSDLERNRCONK 4582
RESULT 14
ADJ83137
ID ADJ83137 standard; protein; 5636 AA.
XX AC ADJ83137;
XX AC
DT 06-MAY-2004 (first entry)
XX DE Human hemimentin protein - SEQ ID 128.
XX KW NOVX; cytostatic; antiarteriosclerotic; antidiabetic; antiasthmatic;
KW antiallergic; antiinflammatory; respiratory; antiarthritic;
KW dermatological; antibacterial; cerebroprotective; vasotropic; cardiant;
KW haemostatic; hypotensive; hepatotropic; neuroprotective; anorectic;
KW nootropic; antitumor; muscular; immunosuppressive; gynaecological;
KW antipsoriatic; endocrine; ophthalmological; osteopathic;
KW antiparkinsonian; anticonvulsant; tranquiliser; analgesic; nephrotropic;
KW antinfertility; antilipemic; cardiomyopathy; atherosclerosis; diabetes;
KW cell signal processing; metabolic pathway; asthma; allergy; emphysema;
KW autoimmune; graft-versus-host; arthritis; cancer; stroke; haemophilia;
KW hemimentin.
XX OS Homo sapiens.
XX PN US2003170630-A1.
XX PD 11-SEP-2003.
XX PF 21-DEC-2001; 2001US-00032189.
XX PR 21-DEC-2000; 2000US-0257495P.
XX PR 22-DEC-2000; 2000US-0258171P.
XX PR 20-FEB-2001; 2001US-0269940P.
XX PR 08-MAR-2001; 2001US-0274192P.

PR 22-MAR-2001; 2001US-0277826P.
PR 29-MAR-2001; 2001US-0279840P.
PR 11-APR-2001; 2001US-0282981P.
PR 13-APR-2001; 2001US-0283656P.
PR 31-JUL-2001; 2001US-0309247P.
PR 10-AUG-2001; 2001US-0311754P.
PR 17-AUG-2001; 2001US-0313331P.
XX (ALSO/) ALSOBROOK J P.
PA (TCHE/) TCHERNEV V T.
PA (LIUX/) LIU X.
PA (SPYT/) SPYTEK K A.
PA (ZERH/) ZERHUSEN B D.
PA (PATT/) PATTURAJAN M.
PA (LEPL/) LEPLEY D M.
PA (BURG/) BURGESS C E.
PA (SHIM/) SHIMKETS R A.
PA (GROS/) GROSSE W M.
PA (SZEK/) SZEKERES E S.
PA (VERN/) VERNET C A M.
PA (LILL/) LI L.
PA (CASM/) CASMAN S J.
PA (BOLD/) BOLDOG F L.
PA (GORM/) GORMAN L.
PA (GANG/) GANGOLLI E A.
PA (FERN/) FERNANDES E R.
PA (RIEG/) RIEGER D K.
PA (EDIN/) EDINGER S R.
PA (GUNT/) GUNTHER E.
PA (MILL/) MILLET I.
PA (SCIO/) SCIORE P.
PA (ELLE/) ELLERMAN K.
PA (MACD/) MACDOUGALL J R.
PA (SMIT/) SMITHSON G.
XX
PI Alsobrook JP, Tchernev VT, Liu X, Spytek KA, Zerhusen BD;
PI Patturajan M, Lepley DM, Burgess CE, Shimkets RA, Grosse WM;
PI Szekeres ES, Vernet CAM, Li L, Casman SJ, Boldog FL, Gorman L;
PI Gangolli EA, Fernandes ER, Rieger DK, Edinger SR, Gunther E;
PI Millet I, Sciore P, Ellerman K, Macdougall JR, Smithson G;
XX WPI; 2003-898249/82.
DR
XX
XX
PT New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,
PT preventing or treating NOVX-associated polypeptide disorder, e.g.
PT cardiomyopathy, atherosclerosis, diabetes, cancer, Parkinson's disease or
PT asthma.
XX
PS Disclosure; SEQ ID NO 128; 263pp; English.
XX
CC The invention relates to a novel isolated NOVX polypeptide. The
CC polypeptide demonstrates cytostatic, antiarteriosclerotic, antidiabetic,
CC antiasthmatic, antiallergic, antiinflammatory, respiratory,
CC antiarthritic, dermatological, antibacterial, cerebroprotective,
CC vasotropic, cardiant, haemostatic, hypotensive, hepatotropic,
CC neuroprotective, anorectic, nootropic, antitumor, muscular,
CC immunosuppressive, gynaecological, antipsoriatic, endocrine,
CC ophthalmological, osteopathic, antiparkinsonian, anticonvulsant,
CC tranquiliser, analgesic, nephrotropic, antinfertility and antilipemic
CC activities. The NOVX polypeptide, nucleic acid or antibody of the
CC invention may be useful for treating or preventing a NOVX-associated
CC disorder, such as cardiomyopathy, atherosclerosis, diabetes or a disorder
CC related to cell signal processing and metabolic pathway modulation.
CC Furthermore, the NOVX polypeptides may be useful for diagnosing, treating
CC or preventing diseases such as asthma, allergies, emphysema, autoimmune
CC disease, graft-versus-host disease, arthritis, cancer, stroke,
CC haemophilia, obesity, Alzheimer's disease and pain. The nucleic acids may
CC be used as hybridisation probes, in chromosome mapping, tissue typing, of
CC preventive medicine or pharmacogenomics. The current sequence is that of
CC a protein of the invention which is related to human NOVX protein.
XX
SQ Sequence 5636 AA;

Query Match		71.6%; Score 2627.8; DB 7; Length 5636;
Best Local Similarity		9.6%; Pred. No. 3.9e-31;
Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;		
QY	1 M-S-----I-----SS--D-----EV-----	8
DB	1 MISWEVHTVFLFALLYSSLAQASQSEIRABEFPAGSTLAFVFDVTGSMYDDLQVVI	60
QY	9 -----NF-LV-----	12
DB	61 EGASKILETSLEKPKPLNFALVPFHPDPEIGPVTTITDPKFKQYELRELYVOGGDCPE	120
QY	13 -----Y-----R--Y-----LQ-----	17
DB	121 MSIGAIAKIALEISLPGSFIVFTDASKDYRLTHEVLQLIQKQSQSVFVLGDCDRTH	180
QY	18 -----E-----SG--F-----S--H--S-----A-----F 26	
DB	181 IGYKVEEIASTSSGVFHLDDKQVNEVLKWVEAVQASKVHLLSTHLEQAVNTWRIPF	240
QY	27 -----T-----FG-----I-----E-----	31
DB	241 DPSLKEVTVLSGSPMIEIRNPLKLIKKGFGHLLNIHNSAKVUNVKEPEAGMTVK	300
QY	32 ---S--H-----I-----S--Q-----	36
DB	301 TSSSGRHVRITGLSTIDFRAGFRKPTLDFKTVSRPVGIGIPTYVLLNTSGISTPARID	360
QY	37 -----S-----NI-----N-----G-----	41
DB	361 LLELLSIGSSSLKTIPIVKYPPHKKPYGIWNISDFVPPNEAFFLKVTGYDKDYLQVSS	420
QY	42 -----A-----L-----VP-----P-----A 47	
DB	421 VSPSSIVPAPKVTMEKTPGYLYQGPQPCSVDSLLPFTLSFRNGVTLGVQYKESA	480
QY	48 -----A-----L-----I-----S-----IIQ-----KG-----	56
DB	481 SVSLDIKVTLSDEGFVECIASVAGTGRAQTFDFVSEPPVQVNNVTVPGERAVLT	540
QY	57 -----L--Q-----P-----Y--V--E-----	61
DB	541 CLIIISAVDNLWQRNDRVRLAEPARITLANLSLELKSVPKFNDAAGEYCHMVSSEGGSS	600
QY	62 -A-----EVS-----NE-----	68
DB	601 AASVFLTVQEPKVTMPKMQSTGSEVSMCSATGYPKPKIAWTVMDFIVGSHRYRM	660
QY	69 ---DGLTF-----D-----G-----R-----P-----I-- 78	
DB	661 TSDGTLFIKNAAPKDAGIYGCLASNSAGTDKQNSTLRYIEAPKLMVVOSELLVALGDIIV	720
QY	79 -E--S-----L--S--LI-----DA-----V-----	87
DB	721 MECKTSGIPPPQWKFGDLPLSTFLIIDLPLGLLKIQETQDLQDAGDYTCVAINAAGR	780
QY	88 -----M-----P-DV-----VQ-----T-----	94
DB	781 ATGKITLDVGGPPVFTQEPADVSMEIGSNVTLPCYQGYPEPTIKWRRLDNMPIFSRPFS	840
QY	95 -----R-----	95
DB	841 VSSISQLRTGALFILNWSADKTYICEAENQFGKIQSETTVTTVTGLVAPLIGISPSVAN	900
QY	96 ---QQ-----A-----Y-----R-D-KL-----AQ-----	105
DB	901 VIEGQQLTFLCTLLAGNPIPERWIKNSAMLLQNPYITVRSGLSHIERVOLQDGGEYTC	960
QY	106 -----OH-----A-----	108
DB	961 VASNAGTNNKTSVVVHVLPITIQHQQLSTIEGIPVTLPCKASGNPKPSVIMWSKKGEL	1020
QY	109 -----A--A-A-----A--A--A-----	114
DB	1021 ISTSSAKFSAGADGSLYVVSPEGESGEYVCTATNTAGYAKRKVQLTVVYRPRVFGDLRG	1080
QY	115 -----A-----A-----	116
DB	1081 LSODKPEISVLAGEEVTLPCEVKSLPPIITWAKETQLISPPSPRHTFLPSSGSMKITET	1140
QY	117 -----ATN-----Q-----Q-GSA--K-----N-----	126
DB	1141 RTSDSGMVLCVATNIAGNVTAQVKNVHVPPKIQRGPKLVQVQGVQVVDIPCAQAQGTPLP	1200
QY	127 -----GE--N-----T--A--N--G-----	133
DB	1201 VITWSKGGSTMLVDGEHVSNDPDTLSIDQATPSDAGIYTCVATNIAGTDETEITLHVQE	1260
QY	134 -----E-----E--N-----GA-----H-----	139
DB	1261 PPTVEDLEPPYNTTPOERVANORIEPPCPAKGTPKPTIKWLHNGRELGTREPGISILEEG	1320
QY	140 T--IA-----N-----N--HT--DM-----	148
DB	1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPPVPIKDKEQVSNVSVLLNQLTNL	1380
QY	149 -----M-----E-----	150
DB	1381 FCEVEGTPSPIIMWKDNVQVTTESSTIQTVNNGKILKLFRTATPEDAGRYSCAKINIAGTS	1440
QY	151 ---VD-----G-----DV--E-----I-----P-- 158	
DB	1441 QKYFNIDVLVPTTIIGTNFPKEVSVVLRDVALQVKGTPPDHFWKDGKPLFGDPN	1500
QY	159 -----SN--K-A-----V-----VL 165	
DB	1501 VELLDRGVHLKNARNDKGRYQCTVSNAAQKQAKDKLTIYIPPSIKGNTTIDISVL	1560
QY	166 -----RG-----	168
DB	1561 INSLIKLECKTRGLPMAITWYKDGQPIIMSSSOALYIDKQVYLHPIRAQVSDSATYCHV	1620
QY	169 -----E-S-----E-----VFI-----C--A-----W-----	177
DB	1621 ANVAGTAESFHVVDVVPPIEMGNLATPLNKQVIAHSLTLECNAAAGNPSPIITLWKDGV	1680
QY	178 -----N-----	182
DB	1681 PVKANDNFRIEAGKKLEIMSAQEDRGOYICVATSVAGEKEIKEYEDVLVPPAIEGGDE	1740
QY	183 -----LL-----VSGS-----G-----D-----S-- 191	
DB	1741 TSVFVIMVNNLELDCHVTGSPPTIMLWKDGLIDERDGFKILLNGRKLVIQAQVSN	1800
QY	192 -----TA-----R-----	196
DB	1801 GLYRCMAANTAGDHKKEPEVTVHVPPTIKSSGLSERVWVYKYPVALQCIANGIPNSITW	1860
QY	197 -----NL--S-----E-----NS--T-----S 204	
DB	1861 LKDDQPWNTAQGNLKIQSSGRVLQIAKTLLEBAGRYTCVATNAAGETQOHIQHVHEPPS	1920
QY	205 ---G-----PTOL-----V-----L-R-----	212
DB	1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLLSGTSMTFLNRGQIIDIES	1980
QY	213 -----H-----	215
DB	1981 AQISDAGIYKCAVINSAGATELFYSLQVHVAPISGSNNMVAVVNNVNRVLECEARGIPA	2040
QY	216 -----R-----E--G-----G--O--D-- 221	
DB	2041 PSLTWLKDGPVSSPSNGLQVLSGGRILALTSTQISDTGRTCYAVNAAGEKQRDIDLVR	2100
QY	222 -VPSN-----	231

Db 2101 YVPPNIMGESQNVSLISQAVELLCSQSDAIPPPPTLTLWKDGHPLLLKPGLSISENRSVLK 2160
QY 232 ---D---W---N---S---EG---TLL--- 240
Db 2161 IEDAQVODTGRYTCEATNAGKTEKKNYNINWPPNIGSGDELQTLTVIEGNLSLICE 2220
QY 241 AT---GS---Y---D---G---F---A--- 249
Db 2221 SSGIPPPNLIWKKGSPVLTDMSGRVRLISGGRQLQISIAEKSDAALYCVASNVAGTAK 2280
QY 250 ---R---I---WTKDG--- 256
Db 2281 KEYNLQVIRPTTNSGSHPEIIVTEKKSISLECEVOGIPPPPTVTKDGHPLIKAKGV 2340
QY 257 ---NL---AS---T---L---GOH--- 265
Db 2341 EILDEGHILQKNIHVSDDTGRYVCVAVNAGMTDKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K---G-PI---FALK---W--- 274
Db 2401 EKNSVLTCSAGIPLSTTWF---KCGWPVLSNSVRILSGGRMLRMQTTMEDAGQYTC 2458
QY 275 ---N---K-K---GN--- 279
Db 2459 VVRNAAGEERKIFGLSVLPPHIVGENTLEDVKKEKQSVTLTCEVTGNVPPEITWHKDG 2518
QY 280 ---F---I---L---S---AGV--- 286
Db 2519 QPLQEDBAHHIISGGREFLOITNVQVPHTGRTYTCCLASSPAGKSRFSNLNVFVSPTIAGVG 2578
QY 287 -D---K---T---I---I-W--- 293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSPPATTIWFKDGTPLESNRNIRILPGRTLOILNAQ 2638
QY 294 -D---A---H---TG---E 299
Db 2639 EDNAGRYCVATNEAGEMIKHYEVKVIPIINKGDLWGPGLSPEKVIKVNNTLLECE 2698
QY 300 A---K---Q--- 303
Db 2699 AYAPASLASWYKDGQPLKDDHVNIAANGHTLIQKEAQISDTGRYTCVASNIAGEDEL 2758
QY 304 F---P-F---A---LDV---D 315
Db 2819 LTSSDKVILPGGRVLIQIPRAKVEDAGRYTCVAVNEAGEDSLOYDVRVLVPIIEGANS 2878
QY 316 ---WQ---SN--- 319
Db 2879 LPBEVTVLNKSALIECLSSGSPAPRNSWQOGQPLEDDHHKFLSGRILQILNTQITD 2938
QY 320 ---NT---FAS---C---S 326
Db 2939 IGRYVCAENTAGSAKYNLNVHVPSPVIGPKSENLTIVVNNFISLTCEVSGFPDPLS 2998
QY 327 ---T---D---MCI--- 331
Db 2999 WLKNXQPIKLTNTLIVPGGRTLQIRAKVSDGGEVTCJAINXAGESKKFSLTVVYVPS 3058
QY 332 ---H---V---C---K---L---GO--- 338
Db 3059 IKDHSESLSVNVREGTSVLECESNAVPPPVITWYKGRMITESTHVEILADGQMLHI 3118
QY 339 ---D---R---P-IK--- 343
Db 3119 KKAESVDTGOYVCRAINVAGDDKNFHLNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 ---T---F---Q---G---H---T--- 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGSKLOIARSOHSDSGNYTCIASNMEGKAQKY 3238

QY 350 ---N-E-V-NA---IKW---D---P---T--- 360
Db 3239 FLISQVPPSVAGABIPSDVSVLGENVELVANGIPTPLIOWLKDGPKPIASGETERIV 3298
QY 361 ---G---NLL---A---S---C---S---D---D---MTL--- 373
Db 3299 SANGSTLNIYGALTSYTGKTCVATNPAGEEDRIFNLNVVYPTTIRGNKDEAKLMTYVD 3358
QY 374 ---K---I-W---SM--- 378
Db 3359 TSINISCRXTGTPTPPQINMLKGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVVTCVASN 3418
QY 379 ---K---Q---DN---CV--- 384
Db 3419 RAGVDNKHYNLQVAFAPNMDNSMGTEITVLKSGTSMACITDGTAPSMAWLRDGOPLG 3478
QY 385 ---H---D---L---OA---H-N---KE 393
Db 3479 LDAHLLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKVLPPHINGSEHEE 3538
QY 394 I---YT---I---K-W--- 399
Db 3539 ISVIVNNPLELTCTIASGIPAPKMTWMKDGRLPQTDQVOTLGGGEVLRISTAQVEDTGRY 3598
QY 400 ---SPTG---P---GT---N--- 407
Db 3599 TCLASSPAGDDDKKEYLVRVHVPPENIAGTDEPRDITVLRNRQVTECKSDAVPPPVITWLR 3658
QY 408 N---P---N---ANLM-LAS-A---S-F--- 420
Db 3659 NGERLOATPRVRLISGRVLOINNADLGDANTYTCVASNIAGKTREFILTIVNVPNIK 3718
QY 421 ---DSTV---RL-W-D-V---DR---G-I--- 433
Db 3719 GPQSLVILLNKSVLESCIAEGVPTPRITWRKQDGLVAGNHARYSILENGLPHIQSAHVT 3778
QY 434 ---C---I---H---TL---T--- 439
Db 3779 TGRYLCMATNAAGTDRRIDLQVHVPPSIAPGTNTMTVINVQVTLTACEATGPKPSINW 3838
QY 440 -K-H---Q-E---P-V---Y--- 446
Db 3839 RKNHLLNVQDNQNSYRLLSSGLSVIISVDDTATVECTVINGAGDDKRTVDTLTQVPP 3898
QY 447 SVA---P-S---P-D--- 453
Db 3899 STADEPTDFLVTKHAPAVITCTASGVPPFSIHWTKNGIRLLPRGDGYRILSSGAIBLAT 3958
QY 454 ---GRY---L---ASG---S- 461
Db 3959 QLNHAGRYTCVARNAAGSAHRVTLHVHBPVIOQOPSELHVILNPNPILLPCEATGTPSP 4018
QY 462 F---D---KCV--- 466
Db 4019 FITWQKEGINVNTSGRHAVALPSGGLQISRAVREDAGTMCVQAQNPAGTALGKILNVQV 4078
QY 467 ---H---I-W--- 469
Db 4079 PPVISPHKEYIAVDKPIITLSCADGLPPPDIITHKDGRAIVESIRQVRLSSGSLQIAF 4138
QY 470 ---N---T---Q--- 472
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTIVHPPRIRSTEGHYTVNENSQAILPCVADGPT 4198
QY 473 ---T---G---A---L-VH 478
Db 4199 PAINKKONVLLANLIGKYTABPYGELILENVLEDGFGVTCVANNAAAGETHVSLTVH 4258
QY 479 ---S---Y---R 481
Db 4259 VLPTTFTELPFDVSLNKGEQLRISCKATGIPKLTWTFNNNIIPAHFDSVNGHSELVIER 4318

Qy 482 -----GT-----G-----G-----IF-----E-----V-C----- 490
 Db 4319 VSKEDSGTYVCTAENSFGVKATGFFVVKPPFKGDPGNSWIEPLGGMALNCEVKGDP 4378
 Qy 491 -----WN-----A-----AGD----- 496
 Db 4379 TPTIQNRKGVDEIHSRIRQLNGSLAIYGTWNEDAGDYTCVATNEAGVVERMSLTLO 4438
 Qy 497 -----K-V-----G-----A-----SAS----- 503
 Db 4439 SPPIITLPEVETINAGGKIILNCQTAGEPQPTITWSRQGHSSWDNRVNVLSNNSLYIA 4498
 Qy 504 -----D-----GSV-----C-V----- 509
 Db 4499 DAQKEDTSEPCVARNLMGSLVRVPVIVQVHGGFSQSAWRACSVTCGKGIQKRSRLCN 4558
 Qy 510 --L-----DL--R-----K 514
 Db 4559 QPLPANGGKPCOGSDLEMRNCQNK 4582

RESULT 15
 ADK60205
 ID ADK60205 standard; protein; 5636 AA.
 XX
 AC ADK60205;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Angiogenesis differentially expressed protein GS-P29.
 XX
 KW vasotropic; antirheumatic; antiarthritic; hypotensive; antiangiinal;
 KW antiinflammatory; cardiant; angiogenesis inhibitor; gene therapy;
 KW angiogenesis; endothelial cell; diagnosis; tumor vascularization;
 KW retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;
 KW ovary hyperstimulation; psoriasis; endometriosis; restenosis;
 KW angioplasty; cicatrization; peripheral vascular disease; hypertension;
 KW vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;
 KW ischemia; angina; myocardial infarction; chronic heart disease;
 KW cardiac congestion; macular degeneration; osteoporosis.
 XX
 OS Homo sapiens.
 XX
 PN FR2836687-A1.
 XX
 PD 05-SEP-2003.
 XX
 PF 11-APR-2002; 2002FR-00004546.
 XX
 PR 04-MAR-2002; 2002FR-00002717.
 XX
 PA (GENE-) GENE SIGNAL.
 PA (ALMA/) AL MAHMOOD S.
 XX
 PI Colin S, Schneider C, Al Mahmood S;
 PI WPI; 2004-013912/02.
 DR N-PSDB; ADK60455.
 XX
 PT Compositions for diagnosing, prognosing and treating angiogenic disorders
 PT including tumor vascularization and heart disease, comprise nucleic acid
 PT or polypeptide differentially expressed in angiogenesis.
 XX
 PS Claim 7; SEQ ID NO 81; 424pp; French.
 XX
 CC The invention relates to a novel pharmaceutical composition active on
 CC angiogenesis comprising an endothelial cell nucleic acid whose expression
 CC is induced by an angiogenic factor and inhibited by an angiostatic agent
 CC or its complement or fragment, a polypeptide sequence encoded by the
 CC nucleic acid or its fragment, a molecule capable of inhibiting expression
 CC of the nucleic acid or a molecule which binds to the polypeptide
 CC sequence. The invention is used to diagnose, prognose or treat an
 CC angiogenic disorder in a mammal, particularly a human. The disorder is

CC particularly tumor vascularization, a retinopathy, rheumatoid arthritis,
 CC Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,
 CC endometriosis associated with neovascularization, restenosis due to
 CC angioplasty, overproduction of tissue due to cicatrization, a peripheral
 CC vascular disease, hypertension, vascular inflammation, Raynaud disease,
 CC aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,
 CC myocardial infarction, chronic heart disease, cardiac congestion or
 CC macular degeneration due to age or osteoporosis. This sequence
 CC corresponds to a protein encoded by a differentially expressed DNA used
 CC in the composition of the invention.
 XX
 SQ Sequence 5636 AA;
 Query Match 71.6%; Score 2627.8; DB 8; Length 5636;
 Best Local Similarity 9.6%; Pred. No. 3.9e-31;
 Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;
 Qy 1 M-S-----I-----SS--D-----EV----- 8
 Db 1 MISWEVVHTVFLFALLYSLAQDASPSQSETRABEPPEGASTLAFVFDVTGSMYDDLQVVI 60
 Qy 9 -----NF-LV----- 12
 Db 61 EGASKILETSLKRPKRPLENFALVPHDPEIGPVTTITDPPKFKQYELRELYVGGGDCPE 120
 Qy 13 -----Y-----R--Y-----LQ----- 17
 Db 121 MSIGAIAKIALEISLPGSFYIVFTDARSKDYRLTHEVLQLIQKQSQVFLVTGDCDRTH 180
 Qy 18 -----E-----SG--F-----S--H--S-----A-----F 26
 Db 181 IGKVVYEETASTSGQVPHLDKKQVNEVLKVEEAVQASKVHLLSTDHLEQAVNTWRIPF 240
 Qy 27 -----T-----FG-----I-----E----- 31
 Db 241 DPSLKEVTVLSGSPSPMIBIRNPLGLIKKGFGHLLHINSAKVNVNKEPEAGMWTVK 300
 Qy 32 ---S--H-----I-----S--Q----- 36
 Db 301 TSSSGRHVRITGLSTIDFRAGFSRXPXTLDFKKTVSRRPVQGIPTYYLLNTSGISTPARID 360
 Qy 37 -----S-----N-----N-----G----- 41
 Db 361 LLELLISGSSSLKTI PVKYYPHRKPYGIWNISDFVPPNEAFFLKVTGYDKDYLFORVSS 420
 Qy 42 -----A-----L-----VP-----P-----A 47
 Db 421 VSFSSIVPDAPKVTWPEKTPGYLQPGQIPCSVDSLPTLSFVRNGVTGLGVQYLKESA 480
 Qy 48 -----A--L-----I-----S-----IIQ-----KG----- 56
 Db 481 SVSLDIKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPVIVQVNNVTVTGERAVLT 540
 Qy 57 -----L--Q-----Y--V--E----- 61
 Db 541 CLIIISAVDYNLTWQRNDRVRLAEPARIRTLANLSLELSKVKENDAGEYHCVSSEGSS 600
 Qy 62 -A-----EVS-----NE----- 68
 Db 601 AASVFLTVQEPKVTVMKQNSFTGGSEVSMCSATGPKPKIAVTNDMFVSGSHRYRM 660
 Qy 69 --DGLTF-----D-----G-----R--P-----I-- 78
 Db 661 TSDGLTFIKNAAPKADAGIYGLCLASNSAGTDKQNSTLRYEAPKLMVMVQSELLVALGDITV 720
 Qy 79 -E--S-----L--S--LI-----DA-----V----- 87
 Db 721 MECKTSGIIPPQVQWFKGDLERPSFTLLIDPLGLLKKIQETQDLQDAGDYTCVAINAGR 780
 Qy 88 -----M-----P-DV-----VQ-----T----- 94
 Db 781 ATGKITLDVSGSPVFTQEPADVSMIEGNSVTLPCYVQGVPEPTIKWRRLDNMFISRPFS 840

QY 95 -----R----- 95
Db 841 VSSISQLRTGALFILNLWASDKGTYYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVAN 900
QY 96 ---QO---A---Y---R-D-KL---AQ--- 105
Db 901 VIEGOQLTLPCTLLAGNPPIERRWIKNSAMLLQNPYITVRSDGLSHIERVOLQDGGEYTC 960
QY 106 -----OH-----A----- 108
Db 961 VASNAGTNNKTTSSVVHVLPTIHOQQILSTIEGIPVLPCKASGNPKPSVIMSKKGEL 1020
QY 109 ---A---A-A---A---A-A--- 114
Db 1021 ISTSSAKFSAGADSLVYVSPGEESGEYVCTATNAGYAKRVQVTVYVPRVFGDLRG 1080
QY 115 ---A---A-----A----- 116
Db 1081 LSQDKPVEISVLAGEEVLTPCEVKSLLPPIITWAKETQLISPFSPRHTFLPSPGSMKITET 1140
QY 117 ---ATN---Q-----Q-GSA-K-----N----- 126
Db 1141 RTSDSGMYLCVATNIAGNTQAVKLVNHPVKIQRGPKHLKVQGVQORVIDPCNAQGTPLP 1200
QY 127 ---GE---N-----T-A-N-G----- 133
Db 1201 VITWSKGSNMLVDGHHVSNPDGTLSDIDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260
QY 134 ---E---E---N-----GA-----H----- 139
Db 1261 PPTVEDLEPPYNTTFQERVANORIEFPCPAKGTPKPTIKWLHNGRELTGREPGISILEEG 1320
QY 140 T---IA---N-----N---HT---DM----- 148
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHPVPIKDEQVSNVSVLLNQLTNL 1380
QY 149 ---M-----E----- 150
Db 1381 FCEVEGTPSIIMWYKDNQVTESSITQTVNNGKILKLFRATPEDAGRYSCKAINAGTS 1440
QY 151 ---VD---G-----DV-E-----I-----P-158
Db 1441 QKYFNIDVLVPTIIGTNFPKEVSVNLNRDVALEQVKGTPFPDIHWFKDGKPLFLGDPN 1500
QY 159 ---SN---K-A-----V-----VL 165
Db 1501 VELLDRGOVLHLKNARNNDKGRYQCTVSNAAQKQAKDIKLTIVIPPSIKGNVTTDISVL 1560
QY 166 ---RG-----H----- 168
Db 1561 INSLIKLECKTRGLPMPAITWYKQGPIMSSQALYIDKQYLHI PRAQVSDSATYTCV 1620
QY 169 ---E-S-----E-----VFI---C-A-----W----- 177
Db 1621 ANVAGTAEKSFHVDVYVPPMIEGNLAPLNKQVVIASHLTLNCAAGNPSPIITLWKDGV 1680
QY 178 ---N-----P-V-SD-182
Db 1681 PVKANDNFRIEAGKKLEINSAQEI DRGQYICVATS VAGEKEIKYEVDVLVPPAIEGGDE 1740
QY 183 ---LL---VSGS---G---D-----S--- 191
Db 1741 TSYFIVMNNLELDCHVTGSPPTIMWLKDGQLIDERGFKILLNGRKLVIQAQVSNT 1800
QY 192 ---TA-----R-----I-W 196
Db 1801 GLYRCMAANTAGDHKEFEVTVHPPTIKSSGLSERVVVVKYKPVALQCIANGIPNFSITW 1860
QY 197 ---NL---S-----E-----NS---T-----S 204
Db 1861 LKDDQPVNTAGNLKIQSSGRVLQIAKTLLDAGRYTCVATNAAGETQQQHIQLHVHPPS 1920
QY 205 ---G-----PTQL---V-----L-R----- 212

Db 1921 LEDAGMNLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLGGSTMTFLNRQIIDIES 1980
QY 213 -----H-----C-----I-- 215
Db 1981 AOISDAGIYKVAINSAGATELFYSLQVHVAPISGSNNMVAVVNNPVRLCEARGIPA 2040
QY 216 -----R-----E-G-----G-Q--D-- 221
Db 2041 PSLTWLKDSPVSSPSFNGVLQVLSGGRIALATSTQISDTGRYTCVAVNAAGEKORDIDLRV 2100
QY 222 -VPSN-----KD-----VT---S-L- 231
Db 2101 YVPPNIMBEEQNVSLISOAVELLQSDAIPPTLTWLDGHPHLLKPKGLSISENSRVLK 2160
QY 232 ---D-----W---N---S-----EG---TLL-- 240
Db 2161 IEQAQVQDTRGYTCATNVAGTEKKNYNNVWPPNIGSGDELDTQLTQVIEGNLISLLCE 2220
QY 241 AT-----GS--Y-D-----G-----F-----A- 249
Db 2221 SSGIPPPNLIWKKGSPVLTDMSGRVRIILSGRQLOISAEKSDAALYSCVASNVAGTAK 2280
QY 250 -----R-----I-----WTXDG----- 256
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSI SLECEVOGIPPTVTWMDKGHPHLLKRAKGV 2340
QY 257 ---NL-AS-----T---L-----GOH----- 265
Db 2341 EILDEGHILQKNIHVSDTGRYCVAVNVAGTMDKKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K-----G-PI-----FALK-W----- 274
Db 2401 EKNSVLTCEASGIPLPSTWPF--KDCWPSVLSNSVRILSGGRMLRLMQTTMEDAQYTC 2458
QY 275 -----N-----K-K-----GN----- 279
Db 2459 VVRNAAEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518
QY 280 -----F-I-----L---S---AGV- 286
Db 2519 QPLQDEAHHIISGGFLOITNVQVPHTRGYTCCLASSPAHGKRSFSLNVFVSPTIAGVG 2578
QY 287 -D---K-T---T-I-----I-W----- 293
Db 2579 SGNPEDVTILNPSPTSLVCEAYSYPATITWFKDGTPLESNRNIIRILPGRTLQILNAQ 2638
QY 294 -D-----A---H-----TG-----E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHVEKVYIPIINKGDLWGLSPKPKVKNVNTLTLECE 2698
QY 300 A-----K-----Q---Q----- 303
Db 2699 AYAI PASASLWYKDGQPLKSDDHVNTAANGHTLQIKEAQISDTGRYTCVASNAGEBELD 2758
QY 304 F-----P-F-----A-----LDV-----D 315
Db 2759 FDNVIOVPFSFQXLWEIGNMLDTRNGEAKDVLIINNPI SLYCETNAAPPTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPIIEGANSND 2878
QY 316 -----WQ-----SN----- 319
Db 2879 LPBEVTVLNKSALIECLSSGSPAPRNSWQKQGPLLEDHKKFLNGRILQILNTQITD 2938
QY 320 -----NT-----FAS--C-----S 326
Db 2939 IGRYVCVAENTAGS AKYFNLVHVPVSVIGPKSENITVVVNNFISITCEVSGPPPDLS 2998
QY 327 -----T-----D---MCI----- 331

Db 2999 WLKXQPIKLTNTLIVPGRTLIIRAKVSDGCVTCIAXNAGESKKKSLFVTVVPPS 3058
QY 332 ---H---V---C---K---L---GO--- 338
Db 3059 IKHDSSESVNVREGTSVSECEBNAVPPVITWYKGRMITESTHVEILADQOMLHI 3118
QY 339 ---D---R---P---IK--- 343
Db 3119 KKAESDGTGOVCRAINVAGRDDKNFHLNVVPPSIEGPPEVIVETISNPVLTCDATG 3178
QY 344 ---T---P---Q---G---H---T--- 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGSKJQIARSQHSDSGNYTCIASNMEGKAQKY 3238
QY 350 ---N---E---V---NA---IKW---D---P---T--- 360
Db 3239 FLSIQVPPSVAGAEIPSDSVLGENVELVNCANGIPTPLIQWLKCKPIASGETERIRV 3298
QY 361 ---G---NLL---A---S---C---S---D---D---MTL--- 373
Db 3299 SANGSTLNIYGALTSDTGKYTCVATNPAGEEDRIFNLNVVYVTPITIRGNKDEAEKLTMYD 3358
QY 374 ---K---I---W---SM--- 378
Db 3359 TSINIECRXTGTPPPQINLKNGLPLSLSSHIRLAAGQVIRIVRAQVSDVAVYTCASN 3418
QY 379 ---K---Q---DN---CV--- 384
Db 3419 RAGVDKNHYNLQVAPFNPMDNSMGTEITVLKGSSTSMACITDGTAPAPSMWLRDQPLG 3478
QY 385 ---H---D---L---QA---H---KE--- 393
Db 3479 LDAHLTVSTHGMVLQLLKAEFTDSGKYTCIASNEAGEVSKHFLKVLPEPHINGSEHEE 3538
QY 394 I---YT---I---K---W--- 399
Db 3539 ISVIVNPLELTCIASGIPAPKMTWMDKGRPLPQDQVOTLGGGEVLRISTAQVEDTRY 3598
QY 400 ---SPTG---P---GT---N--- 407
Db 3599 TCLASSPAGDDKXEYLVRVHVPPNIAAGTDEPRDITVLRNQVTLCKSDAIVPPVITWLR 3658
QY 408 N---P---N---ANLM---LAS---A---S---F--- 420
Db 3659 NGERLQATPRVRLSGRYLIQINNADLGDYANTCVASNIAGKTTFEFLITVNVPPNIG 3718
QY 421 ---DSTV---RL---W---D---V---DR---G---I--- 433
Db 3719 GPQSLVILLNKSTVLECIABGVPTPRITWRKDGAVLAGNHARYSILENGFLHIQSAHVT 3778
QY 434 ---C---I---H---TL---T--- 439
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHVPPSIPAGPTNMTVIVNQVTLACEATGIPKPSINW 3838
QY 440 -K---H---Q---E---P---V---Y--- 446
Db 3839 RKNGHLLNVQDNQNSYRLLSSGSLVIPSVDYATYECTVINGAGDKRTVDLTVOVPP 3898
QY 447 SVA---P---S---P---D--- 453
Db 3899 STADEPTDFLVTKHAPAVITCTASGVFPFSIHWTKNGIRLLPRGDGYRILSSGAIBLAT 3958
QY 454 ---CRY--- 461
Db 3959 QLNHAGRYTCVARNAAGSAHRVTLHVHPPVITQPOPSLHVILNNPILLPCEATGTPSP 4018
QY 462 F--- 466
Db 4019 FITWQEGINVTSGRHAHVLPSSGGLQISRAVREDAGTYMCVAQNPACTALGKILNVQV 4078
QY 467 ---H---I---W--- 469
Db 4079 PPVISPHLKEYIVADKPITLSCEADGLPPPDITWHDGKRAIVESIRQVRLSSGSLQIAF 4138

QY 470 ---N---T---Q--- 472
Db 4139 VQPEDAGHYTCMAANVAGSSSTSKLTIVHVPPIRSTEGHYTVNENSQAILPCVADGIPT 4198
QY 473 ---T---G---A---L---VH--- 478
Db 4199 PAINWKKDNVLLANLLGKYTAEPYGBELILENVVLEDSGFYTCVANNAAAGDTHTVSLTVH 4258
QY 479 ---S---Y---R--- 481
Db 4259 VLPTFTELPGDVSINKGBQLRLSCKATGIPKLTWTFNNIIPAHFDSVNGHSELVIER 4318
QY 482 ---GT---G---G---IF---E---V---C--- 490
Db 4319 VSKEDSGTYVCTAENSVMGFVKATGFVYVKEPPVFKGDYPSNWIEPLGNNAILNCEVKGDP 4378
QY 491 ---WN---A---AGD--- 496
Db 4379 TPTIQWNRKGVDIIEISHRIQLNGSLAIYGTVNEDAGDYTCVATNEAGVVERSMSTLQ 4438
QY 497 ---K---V---G---A---SAS--- 503
Db 4439 SPPIITLBPVETVINAGGKIILNCOATGEBPQPTITWSROGHSISWDDRVNVLNNSLYIA 4498
QY 504 ---D---GSV---C---V--- 509
Db 4499 DAQKEDTSBPECVARNLMGSLVLRVPVIVQVHGGPQSWANRACSVTCGKGIQKRSLCN 4558
QY 510 ---L---DL---R---K--- 514
Db 4559 QPLPANGGKPCQSGDLEMENCONK 4582

Search completed: January 3, 2005, 15:20:14
Job time : 111.333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:10:17 ; Search time 23.6667 Seconds
(without alignments)
1440.315 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 3669

Sequence: 1 MSISSDEVNFLVRYLQESG.....GDKVCASADGSGVCVLDLAK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/PCTRUS_COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2586.1	70.5	4544	1	US-08-469-486-52
2	2586.1	70.5	4544	2	US-08-469-658-52
3	2583.8	70.4	3461	3	US-09-334-220-2
4	2577.9	70.3	4654	3	US-08-476-515A-84
5	2577.8	70.3	4655	3	US-08-652-877-84
6	2577.8	70.3	4655	3	US-08-652-877-88
7	2577.8	70.3	4655	3	US-08-652-877-90
8	2576.8	70.2	4655	3	US-08-652-877-86
9	2560.4	69.8	5405	3	US-08-718-388-9
10	2557.9	69.7	3460	3	US-09-334-220-1
11	2550.8	69.5	15281	2	US-08-471-119A-2
12	2548.4	69.5	4302	3	US-09-052-469-8
13	2548.4	69.5	4302	4	US-08-422-582-8
14	2548.4	69.5	4302	4	US-09-052-262-8
15	2548.4	69.5	4339	3	US-09-052-469-6
16	2548.4	69.5	4339	4	US-08-422-582-6
17	2548.4	69.5	4339	4	US-09-052-262-6
18	2547.1	69.4	3571	4	US-09-911-842A-2
19	2546.3	69.4	4303	2	US-08-460-751-2
20	2546.3	69.4	4303	4	US-09-479-467A-2
21	2545.4	69.4	4302	3	US-08-658-136-5
22	2541.3	69.3	4861	4	US-09-919-497-70
23	2508.2	68.4	3594	4	US-09-911-842A-4
24	2490.2	67.9	3959	2	US-08-970-269A-30
25	2490.2	67.9	3959	3	US-09-407-562-30
26	2486.4	67.8	7257	3	US-09-335-409-5
27	2486.4	67.8	7257	3	US-09-568-102-5

28	2486.4	67.8	7257	3	US-09-567-969-5	Sequence 5, Appli
29	2486.4	67.8	7257	3	US-09-568-480-5	Sequence 5, Appli
30	2486.4	67.8	7257	3	US-09-568-486-5	Sequence 5, Appli
31	2486.4	67.8	7257	3	US-09-568-472-5	Sequence 5, Appli
32	2486.4	67.8	7257	3	US-09-567-899-5	Sequence 5, Appli
33	2485.4	67.7	3727	2	US-08-222-617A-27	Sequence 27, Appli
34	2485.4	67.7	3778	2	US-08-222-617A-2	Sequence 2, Appli
35	2478.9	67.6	4866	4	US-09-424-783-2	Sequence 2, Appli
36	2477.4	67.5	3666	2	US-08-222-617A-12	Sequence 12, Appli
37	2463.8	67.2	3801	2	US-08-822-445-10	Sequence 10, Appli
38	2463.8	67.2	3801	3	US-09-396-540-10	Sequence 10, Appli
39	2458.7	67.0	4928	3	US-09-036-987A-5	Sequence 5, Appli
40	2458.7	67.0	4928	3	US-09-370-700-5	Sequence 5, Appli
41	2458.7	67.0	4928	4	US-09-603-207-5	Sequence 5, Appli
42	2457.1	67.0	3898	3	US-08-750-717-2	Sequence 2, Appli
43	2454.7	66.9	5588	3	US-09-036-987A-6	Sequence 6, Appli
44	2454.7	66.9	5588	3	US-09-370-700-6	Sequence 6, Appli
45	2454.7	66.9	5588	4	US-09-603-207-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-469-486-52
; Sequence 52, Application US/08469486
; Patent No. 5739281
; GENERAL INFORMATION:
; APPLICANT: Thøgersen, Hans Christian
; APPLICANT: Holtet, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: Improved method for the refolding of
; TITLE OF INVENTION: proteins
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,486
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/192,060
; FILING DATE: February 4, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06363/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 542 5070
; TELEFAX: 617 542 8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4544 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-469-486-52

Query Match 70.5%; Score 2586.1; DB 1; Length 4544;
Best Local Similarity 10.0%; Pred. No. 2.4e-32;

Matches 432; Conservative 65; Mismatches 8; Indels 3809; Gaps 387;
Qy 1 MS-----I-----SSDE-----V-----N-F-L-----11
Db 88 MSRLCNGVQDCMDGSDGPHCRELQGNCSRLGCOHCHVPTLDGTCYCNSSFQLOADGKT 147
Qy 12 -----VY-----R-----Y-LQ-----E-----18
Db 148 CKDFDECSVGTCSQLCTNTDGSFICGCVGELLQPNRSCAKNEVPDRPPVLLIANSQ 207
Qy 19 -----SG-----FS-----H-SAF-T-----F 28
Db 208 NILATYLSGAQVSTTPTSTROTAMDFSYANETVCWVHVGDSAAQQLCARMPGLKGF 267
Qy 29 GI-ESH--IS-----Q-----S-N-----I-----NG-----41
Db 268 -VDE-HTINISLSLHVQEMAIDMLTGNFYFVDDIDDRIFVCRNRNGTCVTLDLLELYNP 325
Qy 42 ---AL-----VP-----P-----A- 47
Db 326 KGIALDPAMGVFTFDYGOIPKVERCDMGQNRKLVDSKIVFPHGITLDLVRSLVTWAD 385
Qy 48 A---LI-----SIQKG-L-Q-Y---V-E-----A-EVS-I-----66
Db 386 AYLDYIEWDYEGKGRQTIIO-GILIEHLYGLTVFENLYATNSDANAAQKTSVIRVNR 444
Qy 67 -N-E-----D-GTL-----:-----72
Db 445 FNSTEYQVTVRVKGGALHTYHQRRQPRVRSHACENDQYKPGGCSDCILLANSHKARTC 504
Qy 73 ---F---DG-----R-----R-----PIE-----SL 81
Db 505 RCRSGFLSGDGSKCKPEHEFLVYKGRPGIIRGMDGAKVDPDEHMIENLMNPRAL 564
Qy 82 -----S-LI-----D-----AV-M-----PD 90
Db 565 DPHAETGIYFADTTSYLIQKIDGTERETILKGIHNVGAVDMGMDNLYWTDGPK 624
Qy 91 -V-V-----QTR-----Q-----Y-----99
Db 625 KTSVARLEKAAQTRKTLIEGKTHPRAIIVVDPLNGMWTWDEEDPKDSRRGRLERAMW 684
Qy 100 ---RD---K-----L---A---Q-----Q-----H 107
Db 685 DGSHRDI-FVTSKVLWPNGLSLDIPAGRLYVWDAFYDRIETILLNGTDRIKIVYEGPELNH 744
Qy 108 A-----A-----A-----AA-----111
Db 745 AFGLCGHGNYLFWTEYRSGSVYRLERGVGGAPPTVTLLRSERPPIFEIRMYDAQOQVGT 804
Qy 112 -----AA-----A-A---A-AT---N-----Q-----120
Db 805 NKRVRNCGGSSICLATPGRSQACABEQVLDADGVTCLANPSYPPQCPQGEFACANS 864
Qy 121 ---Q-----G-----S---A---K-----N---G-----127
Db 865 RCIOERWKCDGNDCLNDSDEAPALCHQHTCPSDRKFCENNRICI PNRWLCGDGNDGENSE 924
Qy 128 -E-N-T-A---N-----G-----E-----E---135
Db 925 DESNATCSARTCPPNPFSCASGRCIPTISWTCLDDDCGDRSDESASCAVPTCFPLTQFTC 984
Qy 136 -NG-----A---H---T-----I-----141
Db 985 NNGRCININWRCDNDGNSDEAGCSHSCSTQPKNSGRICI PEHWTCDGNDGCDYS 1044
Qy 142 ---AN---N-----HTD-----M-----M---E---V---151
Db 1045 DETHANCTNQAATPPGCHTDEFQCRLDGLICIPLRWRCDGTDGCMSSDEKSCGTHVC 1104
Qy 152 D-----G-D-----V-----E-I---PS-----NKAV-----163
Db 1105 DPSVKFGCKDSARCIKAWYCDGNDGCDNDECNDESNCEESLACRPPSHPCANNTSVCLPPDK 1164

Qy 164 V-----L-R-----G-H-----E-S-----E-----V- 172
Db 1165 LCDGNDGCGSDGEBGLCDQCGLNCGSCHNSVAPGEGIVCSPLGMELGPDNHTCQIQ 1224
Qy 173 -----F-I-CA---W-----N-----178
Db 1225 SYCAKHLKSCQCDQNKFSVKSCYEGWLEPPDGESCRSLDPKPFIFISNRHEIRIDL 1284
Qy 179 -----P---V-----S-----D-----LL-----184
Db 1285 HKGDYSVLVGLRNTIALDFHLSQSALYWTDDVEDKIYRKGKLLDNGALTSEFVVIQYGLA 1344
Qy 185 -----V-S-----GS-----GD-----S-----T 192
Db 1345 TPGLAVDWIAGNIYVWESNLDOI EVAKLDGTLRTLLAGDIEHPRAIALDPRDGLFWT 1404
Qy 193 ---A---RI-----W-N-----197
Db 1405 DWDASLPRIEASMSGAGRTVHRETGSGGWPNGLTVLYLEKILWIDARSDAIYSARYD 1464
Qy 198 -----LS-----E-----NST-----SG-----205
Db 1465 GSGHMEVLRGHEFLSHPFVAVTLYGGEVYWTDMRTN-TLAKANKWTGHNTVQKNTQPF 1523
Qy 206 ---PT-----Q-----LV-L-R-----H---C-----214
Db 1524 DLQVHPSRQPMAPNCEANGGQPCSHLCLINYNRTVSCACPHLMKMLHKDNTCYEFKK 1583
Qy 215 -----I---RE-----217
Db 1584 FLLYARQMEIRGVLDAPYNYIISFTVPIDNVTVDAREQRYVSDVDTQAIKRAF 1643
Qy 218 ---G-G-----QDVPSNK---D-V-----TS-----LD-----232
Db 1644 INGTGETVVSADLP-NAHGLAVDWVSRNLFWTSYDTNKKQINVARLDGSKFNAVQGLE 1702
Qy 233 -----W---N-S---EG---TLL---A---TG-----S-Y-----245
Db 1703 QPHGLVVHPLRGKLYWTDGDNISMANMDGNSRNTLLFSGQGPVGLAIDPFESKLYMISSG 1762
Qy 246 -----D-GF---A-R---I---W-----T-K-DG-----256
Db 1763 NHTINRCNLGSGLEVIDAMRSQKATALAMGDKLWADQVSEKMGTCCKADGGSVV 1822
Qy 257 ---N---LA-----S---T---L---G 263
Db 1823 LRNSTLLVMHMKVYDESIQLDHHKGTNPCSVNNGDCSQLCLPTSETTRSCMCTAGYSLRSG 1882
Qy 264 Q-----HKG-----P-----I-F-A-----271
Db 1883 QQACEGVGSLYSVHEGIRGIPLDNDKSDALVPVSGTSLAVGIDFHAENDTIYVWDMG 1942
Qy 272 L-----K---W---N---K---KGN-----F-I---L-----282
Db 1943 LSTISRAKEDQWREDVVTNGIRGVEGIADWIAAGNIYWTDOGDFVIEVARLNGSFYVV 2002
Qy 283 -SAGVDK-----T-----T-----I---I-----292
Db 2003 ISQGLDKPRAITVHPEKGYLFWTEWQGYPRIERSRLDGTERRVVLNVNISPNGISVDYQ 2062
Qy 293 ---W-DA-----H-----296
Db 2063 DGKLYWCDARTKIERIDLETGENREVLLSSNNMDMFSVSVFEDFIYWSDRTHANGSIKR 2122
Qy 297 -----TG-----298
Db 2123 GSKDNATDSVPLRTGIGVOLKDIKVFNRDRQKGTNVCAVANGCQQLCLYRGRGORACAC 2182
Qy 299 -----E-A-----K---Q-----Q-----P- 304
Db 2183 AHGMLAEDGASCREYAGYLLYSERTILKSIHLSDERNLNAPVQPFEDPEHMKNVIALAFD 2242

QY 305 -----P--F--H-----S--A----- 309
Db 2243 YRAGTSPGTFNRIFFGDIHFGNIQINDDGSRRITIVENVGSVEGLAYHFGMDTLWTYSY 2302
QY 310 -----P-AL-----D--V-D-----W-Q----- 317
Db 2303 TTSTITRHTVDQTRPGAFERETVITMSGDDHPRAFLDECQNLMTWNEQHPISIMRAA 2362
QY 318 -S-N-----N-----T----- 321
Db 2363 LSGANVLTLEIKDITRPNGLAIDHRAELKLYFSDATLDKIERCEYDGSRYVILKSEPVHP 2422
QY 322 -----P-----A-----SC--S-- 326
Db 2423 FGLAVYGEHI FWTDMVRRARANKHVGNMMLLRVDIPQPMGIIAIVANDTNSCFLSPC 2482
QY 327 -----T-----DM-----CI----- 331
Db 2483 RINGGCDLCLLTHQHVNCSCGRRILODOLTCRAVNSSCRAQDEFECANGECINFSL 2542
QY 332 -----H-----V-----C-K 335
Db 2543 TCDGVPHCKDSEKPSYNSRRCKTFRQCSNRCVSNMLWCNGADDCGDSDEIPCNK 2602
QY 336 -----P-----LG-----Q----- 338
Db 2603 TAGGVGEPRCDGTCIGNSSRCNQFVDCEDASDENMCSATDCSSYFRLGVKGLFQPCER 2662
QY 339 -----D--RP-----I-----K----- 343
Db 2663 TSLCYAPSWVCDGANDCGDYSRDCPGVKRPRCLPNYPACPSGRCIPMSWTCKEDDCE 2722
QY 344 -----T-----P-----Q-----G-----H----- 348
Db 2723 HGEDETHCNKFCSEAFQECQNHRCISKQMLDCGSDDCGSDGDEAHCEGKTCGPSFSFCP 2782
QY 349 -T-----N-----EV--N-A-I-K----- 356
Db 2783 GTHVCVPERWLCDDKDCADGADESAAGLYNSTCDDREFMCQNFQCIKHFVCDHDD 2842
QY 357 -----W-----D-----P--T----- 360
Db 2843 CADGSDSEPECYPTGCPSEFRANCGRCLSSRQWCECDGENDCHDQSDAPKNPHCTSPER 2902
QY 361 -----G-----N-----L----- 363
Db 2903 KGNASSQFLCSCGRCVAEALLCNGQDDCGSDSDRGCHINECLSRKLSGCSQDCEDLKIG 2962
QY 364 -----L--A--S--CS-----D----- 369
Db 2963 FKCRCPGFLKDDGRTCADVDECSITTFPCSQRCINTHSYKCLCVGYAPRGDPSCK 3022
QY 370 -D-----M-----T-LK-----I-W-----SM 378
Db 3023 AVTDEPFLIFANRYLRLNLDGSNTLLKQGLNNAVALDFYRQMIYVTDVTQQGSM 3082
QY 379 -K-----Q-----D--N--C----- 383
Db 3083 IRRMHLNSGVQLHRTGLSNPDGLAVDWVGGNLYWCDKGRDTEVSKLNGAYRTVLVSS 3142
QY 384 -----V-----H----- 385
Db 3143 GLREPRALVVDQVQNGLYWTDGHDHSLIGRIGMDGSSRSVIVDTKITWPNGLTLDVYTER 3202
QY 386 -----D-----L-Q-----AH----- 390
Db 3203 IYWADAREDYIEFASLDGSRHVLVSQDIPHI FALTFLFEDYVYVTDWETKSINRAHKTG 3262
QY 391 -NK-----EI----- 395
Db 3263 TNKTLTISTLRPMDLHVHALRQDPVPHPCVNNNGGCSNLCLLSPGGGHKACPTNFY 3322
QY 396 -----T-----IK--W-----S-----P----- 401

RESULT 2

Db 3323 LGSRGRTCVSNCTASQFVCKNDKCIFFMWKCDTEDDCGHSDPDPBPCFKRPGQFQS 3382
QY 402 TG-----P-----G-----TN-----N----- 408
Db 3383 TGICTNPAPICDGDNDQDNDSEANCDSHVCLPSPQKCTNTNRKICIFRCNGQDNGDG 3442
QY 409 -----PN-----ANL--M----- 414
Db 3443 EDERCPEVTCAPNQFQCSITKRCIPRVWVCDRNDNDVDSDEPANCCTQMTGCVDFRCK 3502
QY 415 -----L-A----- 416
Db 3503 DSGRCIPARWKCDGEDCGDSEPKCECDERTCEPYQFRCKNNRCVPGRWQCDYDNCG 3562
QY 417 -----S-----A--S-----E----- 420
Db 3563 DNSDBESCTPRCSSESEFCANGRCIAGRWKCDGDDHDCADGSDKDEKCTPRCDMDQFOCKS 3622
QY 421 -----D-S-----T-VR-----L-W----- 427
Db 3623 GHCIPLRWCADADCMGDSDEACGTGVRTCPLDEFQCNNTLCKPLAMKCDGEDDCGDN 3682
QY 428 -D-----V-----DRGIC-I-----HTLTK 440
Db 3683 SDENPEECARFVCPNRPFRCKNDR-VCLWIGRQCDGTDNCGDGTDEDECEPPTAHT-T- 3739
QY 441 H-----Q-----E-----P-----V 445
Db 3740 HCKDKKBFCLRNQRCUSSSLRCNMFDDCGDSEDEDCSIDPKLTSCATNASICGDEARCV 3799
QY 446 -----Y--S--V-----A-F----- 450
Db 3800 RTEKAAACACRSGFHTVPGQCGDINECLRFGTCSQLCNNTKGGHLCSCARFMKTHNT 3859
QY 451 -----S-----P-----D-----GR----- 455
Db 3860 CRAEGSEYQVLYIADNEIRSLPPGPHPSAYEQAFQGDSESVRIDAMDVHVKAGRVYTNW 3919
QY 456 -----Y--L--A--S-----G-----F----- 462
Db 3920 HTGTISYRSLPPAAPTTSNRHRRQIDRGVTHLNLISGLKMPRGIAIDWAGNVYWTDSGR 3979
QY 463 D-----K-C-----V--H-I-----W-N-----T- 471
Db 3980 DVIEVAQMGKGNKTLISGWDIDEPHAI VVDPLRGTYWSDGNHNPXIETAAMDGLRETL 4039
QY 472 -----Q--TGAL-V--HS--Y-----R-GT-----G----- 484
Db 4040 VQDNQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPIVAADSKRGLSHPSIDVF 4098
QY 485 -----G-----I--F-----EV-----C-W 491
Db 4099 EDYIYGVTVYNNRVFKHKFGHSPVLNLTGGLSHASDVVLYHQHKOPEVTNPCDRKKEW 4158
QY 492 -----N-----A-----A----- 494
Db 4159 LCLLSPSGPVCTCPNGKRLDNGTCVVPSPPTPPDAPRPGTCNLQCFNGSGCFLNARQP 4218
QY 495 -----GDK-----V--G--A-- 500
Db 4219 KRCQPRYTGDKCELDCQWEHCNNGGTCAASPSGMTCPCTGTGTGPKCTQQVCAGYCAN 4278
QY 501 -----S-----ASDCS----- 506
Db 4279 NSTCTVNOGNQPCRLCPLGLGDRCOYRQCSGYCENFGTCQMAADGSRQCRCTAYEGSR 4338
QY 507 --V-----CVL-----DL-----R-----K 514
Db 4339 CEVNCSCRLEGACVVKQSGDVTCTDGRVAPSLCTCVGHCSNGSGSCTMNSK 4392

```

US-08-469-658-52
; Sequence 52, Application US/08469658
; Patent No. 5917018
; GENERAL INFORMATION:
; APPLICANT: Th egersen, Hans Christian
; APPLICANT: Holtet, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOOLDING OF
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,658
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIORITY INFORMATION:
; APPLICATION NUMBER: 08/192,060
; FILING DATE: February 4, 1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06363/002002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 542 5070
; TELEFAX: 617 542 8906
; TELETYPE: 200154
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4544 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-469-658-52

Query Match 70.5%; Score 2586.1; DB 2; Length 4544;
Best Local Similarity 10.0%; Pred. No. 2.4e-32;
Matches 432; Conservative 65; Mismatches 8; Indels 3809; Gaps 387;

QY 1 MS-----I-----SSDS-----V-----N--P-L-----11
|||
|||
|||
Db 88 MSRLCNGVQCDMGDEGPHRCLQNCRLGCGHCVPLDGTGTCVCSNFFOLQADGKT 147
|||
|||
|||
QY 12 -----VY-----R-----Y-IQ-----E-----18
|||
|||
|||
Db 148 CKDFDECVYGTCSQLCTNTDGSFICGCEGYLLQPDNRCSKAKNEFVDRPPVLLIANSQ 207
|||
|||
|||
QY 19 -----SG-----FS-----H--SAF-T-----F 28
|||
|||
|||
Db 208 NILATYLSGAQVSTIPTSTROTMTDMDFSVANETVCWVHVGDSAAQTOLKCARMPGLKGF 267
|||
|||
|||
QY 29 GI-ESH-----IS-----Q-----S-N-----I-----NG-----41
|||
|||
|||
Db 268 -VDE-HTINISLSLHVEQMAIDWLTGNYFVDDIDRIFCVCRNGTCTVLLDLELYNP 325
|||
|||
|||
QY 42 ---AL-----VP-----P-----A-47
|||
|||
|||
Db 326 KGIALDPAKGVFTDYGQPKVRCRMDGQNTKLVDSKIVFPHGITLVLRLVYVWAD 385
|||
|||
|||
QY 48 A---LI-----SIQKG-L-Q--Y---V-E-----A-EVS-I---66
|||
|||
|||

```

Db 1465 GSGHMEVLRGHEFLSHPFVATLYGGEVYWTMRN-TLAKANKWTGHNVTVQRTNTQPF 1523
QY 206 -PT- -Q- -LV-L-R- -H- -C- - 214
Db 1524 DLOVHPSRQPMAPNCEANGGQPCSHLCLINYNRTVSCAPHLMKLKDNTTCYEFKK 1583
QY 215 -G-G- -QVPSNK- -D-V- -TS- -LD- - 232
Db 1584 FLLYARQMEIRGVLDAPYNYIISFTVPIDNVTLVDAREQRYVWSVDRVTOAIKRAF 1643
QY 218 -G-G- -QVPSNK- -D-V- -TS- -LD- - 232
Db 1644 INGTGVETVSADLP-NAHGLAVDWVSRNLFMTSYDTNKKQINVARLDGSGFKNAVVGLE 1702
QY 233 -W- -N-S- -EG- -TLL- -A- -TG- -S- -Y- - 245
Db 1703 QPHGLVHPLRGKLYWTGDNISMANMDSNRLLFSGQKGPVGLAIDFPESKLYWISSG 1762
QY 246 -D- -GF- -A- -R- -I- -W- -T- -K- -DG- - 256
Db 1763 NHTINRCNLGSGLEVIDAMRSQKATALAIMGDKLWADQVSEKMGTCRADGSGSVV 1822
QY 257 -N- -LA- - - -S- -T- -L- -G- 263
Db 1823 LRNSTLVMHMKYDESIQLDHKGTNPVSNNGDCSQCCLPTSETTRSCMCTAGYSLRSG 1882
QY 264 Q- - -HKG- -P- -I- -F- -A- - 271
Db 1883 QQACEGVGFLYSVHEGIRGIPLDNPKSDALVPVSGTSLAVGIDFHAENDTIYVDMG 1942
QY 272 L- - -K- -W- -N- -K- -KGN- -F- -I- -L- - 282
Db 1943 LSTISRAKRDQWREDVWNGIRVEGIADVMAGNIYWTDOGFVIEVARLNGSPRYV 2002
QY 283 -SAGVDK- -T- - -I- - 292
Db 2003 ISQGLDKPRAITVHPEKGYLFWEMQYPRIERSLUDGTERVVLNVNVSISWENGISVDYQ 2062
QY 293 -W- -DA- - -H- - 296
Db 2063 DGKLYCDARTDKIERIDLETGENREVLSSNNMDFSVFEDFIYWSDRTHANGSIKR 2122
QY 297 -TG- - - - 298
Db 2123 GSKDNATDSVPLRTGIGVOLKDIKVFNRDRKGTNVCAVANGCQQLCYRGRGORACAC 2182
QY 299 -E- -A- -K- - -Q- - -F- - 304
Db 2183 AHGMLAEDGASCREYAGYLLYSERTILKSIHLSDERNLNAPVQPFDPPEHMKRNIALAFD 2242
QY 305 -P- -F- -H- -S- -A- - 309
Db 2243 YRAGTSPGTPNRIFFSDIHFGNIQINDGSRITIVENVGSVEGLAYHRGWDTLWTSY 2302
QY 310 -P- -AL- -D- -V- -D- -W- -Q- - 317
Db 2303 TTSTITRHTVDQTRPCAFARETIVITMSGDHHPRAFLVLDQCNLMFWTNWNEQHPISMRAA 2362
QY 318 -S- -N- -N- -T- - 321
Db 2363 LSGANVLTLEIKDIRTPNGLAIDHRAEKLYFSDATLDKIERCEYDGHSHRYVILKSEPVHP 2422
QY 322 -F- - -A- -SC- -S- - 326
Db 2423 FGLAVYGEHIFWTDWVRRAVQRANKHVGNMKNLLRVDIFQOPMGIIIVANDTNSCELSPC 2482
QY 327 -T- - -DM- -CI- - 331
Db 2483 RINNGCQDLCLLTHGHVNCRCGRILQDDLTCEAVNSSCRAQDEFECEANGECINFSL 2542
QY 332 -H- - -V- -C- -K- 335
Db 2543 TCDGVPHCKDKSEKPSYCNRRCKTFRQCSNGRCVSNMLWNGADDGCGDSDEIPCNK 2602

QY 336 - - -LG- -O- - 338
Db 2603 TAGGVGEPRCRDGTCTIGNSSRCNQFVDCEDASDEMNCSATDCSSYFRLGVKGVLPQPCER 2662
QY 339 -D- - -RP- -I- -K- - 343
Db 2663 TSLCYAPSVWCDGANDCGDYSDERDCPGVKRPRCLNYPACPSGRICIPMSWTCKEDDCE 2722
QY 344 -T- - -F- -Q- -G- -H- - 348
Db 2723 HGEDETHCNKFCSEAQFECQNHRCISKOMLDCGSDDCGDSDEAAHCEGKTCGSPSFSCP 2782
QY 349 - - -N- -N- -A- -I- -K- - 356
Db 2783 GTHVCVPERWLCDGDKDCADGADESIAGCLYNSTCDDREFMCONQCIPKHFVCDHDD 2842
QY 357 - - -W- - -D- -P- -T- - 360
Db 2843 CADGSDSEPECYPTCGPSEFRCANGRCLSSRQWECDEGNDCHDQSDAPKPNPHCTSPH 2902
QY 361 - - -G- -N- -L- - 363
Db 2903 KCVASSQFLCSSGRCAEALLCNGDDCGSDSDERGHINECLSRKLSGCSQDCEDLKIG 2962
QY 364 -L- - -A- -S- -CS- -D- - 369
Db 2963 FKCRCPGFRLLKDDGTCAVDDECSTTFPCSORCINTHGSYKCLCVGYAPRGDHPHCK 3022
QY 370 -D- - -M- -T- -LK- -I- -W- -SM 378
Db 3023 AVTDEEPFLIFANRYVLRKLNLDGSGNYTLKQGLNNAVALDQYREQMIYVTDVTTQSGM 3082
QY 379 -K- - -Q- -D- -N- -C- - 383
Db 3083 IRRMHLNGSNVQLHRTGLSNPDGLAVDWVGNLYWCDKGRDTIEVSKLNGAYRTVLVSS 3142
QY 384 -V- - -H- - 385
Db 3143 GLREPRALVVDVQNGYLYWTDGHSLSIGRIGMDGSSRSVIVDTKITWPNGLTLDYVTER 3202
QY 386 -D- - -L- -Q- -AH- - 390
Db 3203 IYWADAREYIEFASLDGSRNRHVLSDQIPHIFALTDFEDYVYVWTDWETKSINRAHKTG 3262
QY 391 -NK- - -EI- - -Y 395
Db 3263 TNKTLILLSTLHRPMDLHVPHALRQPDVPHPCVKNNGGCSNLCIASPGGGHKACPTNFY 3322
QY 396 -T- - -IK- -W- -S- -P- - 401
Db 3323 LGSGRTCVSNCTASQFVCKNDKCIFFWVKCDTEDDCGHSDDEPPDCPFKCPGQFQCS 3382
QY 402 TG- -P- -G- -TN- -N- - 408
Db 3383 TGICTNPAFICDGDNDQDNDEANCDIHVCLPSQFKCTNTNRCIPGIFRCNGQDNCGDG 3442
QY 409 - - -PN- - -ANL- -M- - 414
Db 3443 EDERDCPEVTCAPOFQCSITKRCIPRVWVCDRNDVCDGSDPANCTQTCGVDEFRCK 3502
QY 415 -L- -A- - 416
Db 3503 DSGRCIPARWKCDGDDCGDSDPEKCECDERTCEPYQFRCKNNRCVPGRWQCDYDNDG 3562
QY 417 -S- - -A- -S- -F- - 420
Db 3563 DNSDEESCTPRCSSESEFSCANGRCIAGRWKCDGHDGDCADGSDKDXCTPRCMDQFQCKS 3622
QY 421 -D- -S- -T- -VR- -L- -W- - 427
Db 3623 GHCIPLRWRCDADACMDGSDSEACGTGVRTCPLDQFCNNTLCKPLAWKCDGEDDCGDN 3682

QY 428 D-V-----V-----DRGIC-I-----HTLTK 440
 Db 3683 SENPEECARFCPPNRPFRCKNDR-VCLWIGRQCDDTNCDDGDEDCBPPTAHT-T- 3739
 QY 441 H-----Q-----E-----P-----V 445
 Db 3740 HCKDKKEFLCRNQRCLSSSLRCNMFDDCGDGDDEDCSIDPKLTSCATNASICGDEARCV 3799
 QY 446 -----Y-----S-----V-----A-F----- 450
 Db 3800 RTEKAAACACRSGFTVPQPGQDINECLRFGTCSQLCNNTKGHLCSCAENFWKTHNT 3859
 QY 451 -----S-----P-----D-----GR----- 455
 Db 3860 CKAEGSEQVLYIADNNEIRSLFPGHPHSAVEQAQGDSESVRIAMDVHVKAGVYWTNW 3919
 QY 456 -----Y-L-A-S-----G-----S-----F----- 462
 Db 3920 HTGTISYRSLPPAAPPTTSNRHRRQIDRGVTHLNISGLKMPRGTAIDWVAGNVYWTDSGR 3979
 QY 463 D-----K-C-----V-H-I-----W-N-----T- 471
 Db 3980 DVIEVAQKGENRKTLSGMDIPEHAIIVDPRLRGTMYWSDWGNHPKIETAAMDGTLRETL 4039
 QY 472 -----Q-TGAL-V-HS-Y-----R-GT-----G----- 484
 Db 4040 VODNIQWFTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPDIVAADSKRGLSHPPFSDVF 4098
 QY 485 -----G-----I-F-----EV-----C-W 491
 Db 4099 EDYIYGVTYINNRVFKIHKFGHSPVLNLTGGLSHASDVVLYHQHQPVEVTPNCPDRKKCEW 4158
 QY 492 -----N-----A-----A----- 494
 Db 4159 LCLLSPGPVCTCNGKRLDNGTCVPVSPPTPPDAPRPGTCNLQCPNGSGSCTFLNARRQP 4218
 QY 495 -----GDK-----V-G-A- 500
 Db 4219 KCRCQPRYTGDKCBELDQWEHCRNGGTCAASPSPGMPTRCPTGFTGPKCTQQVCAGYCAN 4278
 QY 501 -----S-----ASDGS----- 506
 Db 4279 NSTVTNQGNOQRCRLPGFLGRQYRCQSGYCNENFGTCMAADGSRQCRCTAYFEGSR 4338
 QY 507 -V-----CVL-----DL-----R-----K 514
 Db 4339 CEVVKCSRCLGACVNVKQSGDYTCNCTDGRVAPSLCTCVGHCSNGSGSCTMNSK 4392

RESULT 3

US-09-334-220-2
 ; Sequence 2, Application US/09334220
 ; Patent No. 6323177
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Jude's Children's Research Hospital
 ; APPLICANT: Curran, Thomas
 ; APPLICANT: D'Arcangelo, Gabriella
 ; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW
 ; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND
 ; TITLE OF INVENTION: THERAPIES
 ; FILE REFERENCE: 2427/05704
 ; CURRENT APPLICATION NUMBER: US/09/334,220
 ; CURRENT FILING DATE: 1999-06-16
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 3461
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-09-334-220-2

Query Match 70.4%; Score 2583.8; DB 3; Length 3461;
 Best Local Similarity 11.7%; Pred. No. 6.2e-33;

Matches 392; Conservative 93; Mismatches 16; Indels 2842; Gaps 359;
 QY 1 M-----S-I-----S-----D-----EV----- 8
 Db 1 MERGCWAPRALVLAVLLLLATLRARATGYPPRFPFLCTHHELGDEGBOGEVULISL 60
 QY 9 ---N-----F-LV--YRY---LQ--ES---GF--S-H 23
 Db 61 HIAGNPYYVPGQBYHVTISTSTFFDGLLVGL-YTSTSIQSSQSIGSSAFGFGIMSDH 119
 QY 24 ---S-A-----F-----T-----F-----G-I----- 30
 Db 120 QFNGQWCVSVASHVSHLPNTNLSFWIAPPAGTCVNFMATATHRGQVIFKDALAQOLC 179
 QY 31 ---E---SH-----I-----S-O---SNI-----N-----GAL----- 43
 Db 180 EQCAPTEATAYSHLAETHSDVILRDDFDYSYQLELNPNIWVECSNCEMGEQCQTTHMGN 239
 QY 44 -V---P-P-----A-AL--I-----SI-----IQ-- 54
 Db 240 AVTFCEPYGPRELTTTCLNNTTASVLQFSIGSGSCRFSDPSITVYAKUNTADMIQLE 299
 QY 55 -----KG---LQY---V-EA-E-----VSIN----- 67
 Db 300 KIRAPNSVTVIHLYPEEAKGESVQFQWKQDSLVRGEVACWALDNILV-INSARE 358
 QY 68 ---ED---GT-LF-----DG-----R----- 76
 Db 359 VVLEDNLDPDVTGNWLFPPGATVHSCQSDGNSIYFHGNEGSEFNATTROVDLSTEDIQ 418
 QY 77 -----P-I-----ES-LSLI-----DA----- 86
 Db 419 EQWSEFEFSQPTGWDILGAVGADCGTVESGLSVFLKDGERRKLCPTPYMDTTGYGNLRFY 478
 QY 87 -VM-----P-----D-----VV-----QT----- 94
 Db 479 FVWGIGCDPGVSHENDIILYAKIEGRKEHIALDTLYSSYKVPVSLVSVVINPELQTPATK 538
 QY 95 ---RQ---QAY-R---D-----K-LAQ-----Q----- 106
 Db 539 FCLRQKHOGYRNVRNVAVDFFHVLVPLPSTMSHMIQFSINLGGCTHQPQNSVLSFSTNH 598
 QY 107 ---HA-----A-----AAA-AA-----AA-----AT-- 118
 Db 599 GRSWSLLHTECLPEICAGPHLPHSTVYSENYSYGNWRITIPLPNAALTRDTRWRQTP 658
 QY 119 -----N-----Q-----Q-----GS 123
 Db 659 ILGNMWAIDNVYIGPCLKFCSCGRQCTRHGCKCDPGFSGPACEMASQTFPMFISESGS 718
 QY 124 AK---N-----G-E-----N----- 129
 Db 719 ARLSSYHNPFYSIRGAEVSGCGVLASGKALVFNKDRRLQTLITFLDSSQSREFLOFTLRIG 778
 QY 130 ---T-A-----NG--E-E--N-----G- 137
 Db 779 SKSVLSTCRAPQPCGEGVLLHVSYDNGITWKLLEHYVYNHEPRIISVELPDDARQGI 838
 QY 138 ---A-H-----T-----I-----AN-----N-----H- 145
 Db 839 QFRWQPYHSSQGEDVWAIDEIVMTSVLFNSISLDFTLNVEVTQSLGFLGNVQPCGHD 898
 QY 146 ---T-D---M-----MEVDG-----D-V-E----- 156
 Db 899 WTLCTGDSKLASSRMYVETOSMQI-GASYMIQPSLVNMGCGQKYTPHMDNVKLEYSANH 957
 QY 157 ---I-----PSN---K-----A-----VVL-----R----- 166
 Db 958 GLTWHLVQEECLPSPSCQEFTSASIVHASEFTQWRVTVVLPOKWSGATFRWSQSY 1017
 QY 167 -----GH-----E-SE-----V---F--- 173
 Db 1018 TAQDEWALDNIIYGOCCPNMCSGHGSDHGVCRCDQGYQGTGTECHPEALPSTIMSDFENP 1077

QY 174 -----I-----C-----A-----W----- 177
Db 1078 SSWESDQWQVIGGVVKEQCGVSSGSSLYFSKAGKROLVSWDLDTSWDFVQFYQI 1137
QY 178 -----N-P-----V-----SDL-----LV-----SG----- 187
Db 1138 GGSAAACNPKDPSREEGILLOYSNNGIGTOMHLLAEMVFSDFSKPRFVYVLELPAAGKTPCTR 1197
QY 188 -----SG-----D-----S-----T-----A-----R-IR- 196
Db 1198 FRWKPVSCEGYDQWAVDIIILSEKOKOVIPIVNPPTLPONFYKPAFDYPNQMOSVWL 1257
QY 197 -----N-LSN-S-----T-----S-G-----P-----T-QL----- 209
Db 1258 MLANEGMAKNDSFCAITPSAMVFGKSDGRFAVTRDLTLKPGVQLQKLNIGTSQFSST 1317
QY 210 -----V-LR-H-----C-----IR-EG-----GQD-----V 222
Db 1318 APVLLQYSHDAGMSWFLLEKGCFFPASAAGCEGNSRELSEPTVYTG-DPEENTRITIAI 1376
QY 223 P-----S-N-----K-----D----- 227
Db 1377 PRSLASKTRFRWIOESSOKNPPFGLDGVYISEPCPSYCSHGDCISGVCFDLGYTA 1436
QY 228 -----V-T-----S-LD-----W-----NSEG-TL-----L-----A- 241
Db 1437 AQGTCVSNTPNHSEMFDPRPEGLKPLWYKITGGQVGTGCTLN-DGRSLFYNGLGKREAR 1495
QY 242 -----T-----GS-----Y-D-G-----F 248
Db 1496 TVPLDTRNISLVQFYQIGSKTSGITVTPRAREYGLVQVSNNDGILWHLLRELFMSF 1555
QY 249 -----A-RW-----T-----K-DG- 256
Db 1556 LEQIISIDLPREAKTATAPR-WMOPOHKGHSAQWALGVILGVNDSQTGFQDKLDGS 1614
QY 257 -----N-----LAS-T-L-----G-Q-----H-----KG- 267
Db 1615 IDLOANWYRLOGQVQDIDCL-SMDTALIFTENIGNPRYAETWDFHVSSESPLOWENMGC 1673
QY 268 -----P-P-A-----L-----K-W-----NKK-----G----- 278
Db 1674 SKP-FSGAHGILQYSLNKGQWLTTECVPPTIGCVHYTESSTYTSERFQNWRRVTV 1732
QY 279 -----N-----F-I-----LSAGYDKT-TI-----I-----W-----DA----- 295
Db 1733 LPLATNSPRTRFRWIOQNTYTVGAD-SWAIDNVILASGCPMWCGRGICSDRGVCDRGFG 1791
QY 296 -----H-----T-----GE----- 299
Db 1792 GPFCVPVPLPSILKDDPFGNLHPDLWPEVYGAERGNLNGETIKSGTCLIFKGEGLRMLI 1851
QY 300 -----AK-Q-Q-----P-----F-----H-----S----- 308
Db 1852 SRDLDCNTMYQVFSRLFIAGKTPERSHSLILQFSVSGVTWHLMBDEFYPPQTTSLFIN 1911
QY 309 AP-----A-----L-----D-----V-----D-W 316
Db 1912 VPLPYGAQTNATRFLWQPYNKGKBEIWIIDDFIIDGNLNNPVLILLDTDFGPNEDNW 1971
QY 317 -----Q-S-N-NT----- 321
Db 1972 FFPYGGNIGLYCPSYKGAPEEDSAMVFSVNEGEHSITTRDLSVNETIIQFEINVGCS 2031
QY 322 -----F-A-----S-CST----- 327
Db 2032 TDSSADPVRLEFSRDFGATWHLILLPLCYHSSSLVSLCSTEHSPSTVYAGTTQWRRE 2091
QY 328 -----D-----MC-IH-----VC 334
Db 2092 VWHFGKLHLCGVRFRWYQFYFAGSQPVTWIADNVYIGQCCEMCGHGSCTINGTKCIC 2151

QY 335 -----K-L-----GQ-----DR-----P-----I-KT-----F-Q-G- 347
Db 2152 DPOYSGPTCKISTKNPDLKDDPEGQLESDDRFLMSGGKPSRKCGILSSGNLFFNEDGL 2211
QY 348 -----H-----T-N-----EV----- 352
Db 2212 RMLVTRDLDLSHARFVQFMRLGCGKGVPPRSPQVLLQYSLNGGLSWSLLQBLFLFSNSS 2271
QY 353 NA-----I-----K-----W----- 357
Db 2272 NVGRYTALEMPKARSSTRLRWQPSENGHFYSPVVIDOILIGGNISGNTVLEDDPSTL 2331
QY 358 D-----P-----TG-----N-----LL-----ASC-S-D 369
Db 2332 DSRKLLHPGCTKMPVCGSTGDALVFEKASTRYVVTDDIADVNEGSFLQIDFAASCSTVD 2391
QY 370 -----DM-----T-----L-KI-----W----- 376
Db 2392 SCVAIELEYSVDLGLSMHPLVRDCLPTNVECSRYHLQRLVSDTFNKTWTRITLPLPSYTR 2451
QY 377 S-----M-----KQ-----DN-----CV-----H-----D- 386
Db 2452 SQATRFWRHQPAPFDKQOTWADNVYIGDCLDMCSHGRCVQSGCVDEQWGLYCDPE 2511
QY 387 -----L-----Q-----A-H----- 390
Db 2512 ETSPLTQLKDNFRAPSNQNLTVSGKLSLTVCGAVASGLALHFGSGCSRLLVTLVDLNL 2571
QY 391 NKE-I-----Y-----T-----IKW-----S-PTG-----P 404
Db 2572 NAEFIOFYMYGLITPFSNRNOGVLLSEYVNGGITWNLMEIFYDQYSKP-GFVNILLPP 2630
QY 405 -----CT-----NNP-----N-A-----NLMLA-SA-----SF-----D 421
Db 2631 DAKBIATRFWRWQ-PRHDGLDQNDWADNVILSGSADQRTVMDLTFTSSAPVPOHRSAPD 2689
QY 422 S-TV-R-----L-----W-----DVD-----R-----GI-C-IH-----TL-TKH- 441
Db 2690 AGPVGRIAFEMFLEDKTSVNEENWLPD-DCTVERPCDSPGVMLCGSHDGRVAVT-HD 2747
QY 442 -----Q-----E-P-----V-YSAF-----S-----P-D----- 453
Db 2748 LPTENWIMQFKISVGCKVPEKIAQNIHVQFSTDFGVSWSYLVLPQCLPADPKCSGSVQ 2807
QY 454 -----G-R-Y-L-AS-G-----S-----F-----D- 463
Db 2808 PSVFPTEGWKRIITYPLPESLTGNVFRFPYQKYSVQWADNFYLPGLGCLDNGGHCDC 2867
QY 464 -K-CV-----H-----I-----W-----NT- 471
Db 2868 LKEQCICDPGSGPNCYLTHSLKTLKRFPSDBEIKPDLWMSLEGSGTCTECGVLAEANTA 2927
QY 472 -----Q-T-----GA-L-----V-----H-----S----- 479
Db 2928 LYFGSGTVAQAITQDLDLRGAKFLQYWGRIQSGENNNMTSHRPPVCRKEGVLLDFSTGGIT 2987
QY 480 -----Y-R-----G-----T----- 483
Db 2988 WTLLEHMDFOKYISVRHDYILLPEGALTWTRLRWQPPFVSNGLVVSVERAQWALDNI 3047
QY 484 --GG-I-----P-----E-----V-----C-----W-----NA 493
Db 3048 LIGGAEINPSQLVDTDFDESGSHEENWSFYPNAVRTAGFCGNFSPHLYWPKKKDKTHNA 3107
QY 494 -----A-----GD-----K----- 497
Db 3108 LSSRELIQGYMMQPKIVVGCEATSGDLHSMLEYTKDARSWSQLVQTCLPSSNS 3167
QY 498 VG-----A-----SA-----S-D- 504
Db 3168 ICGSPQFHEATYNAVNSSSWKRTITIQLDHVSSATQFRWIKGEEETKQSWAIDHVV 3227
QY 505 -G-----S-----VC----- 508

Db 1140 EKNCNSTETCQPSQFNCNHRICIDLSFVCDGDKDCVDGSDVGVCLNCTASQFKCASGDK 1199
QY 163 ---V---VL---R-G-H---E---169
Db 1200 CIGVTRCDGVPCSDNSDEAGCPTRPPGMCHSDEFQCEQEDGICINFWECDGHPDCLYG 1259
QY 170 S-E-V---F---IC---AW---N---P---179
Db 1260 SDEHNACVPTCPSSYFHCNCGNCHRAWLCDRDNDCGMSDEKDCPTQFPFCPSQWQC 1319
QY 180 ---V---SD---182
Db 1320 LGHNICVLSVWCDGIFDCPNGTDESPLCNGNSCDFNGGCTHECVQBPFGAKCLCPLGF 1379
QY 183 LL---V---S---GS---G---189
Db 1380 LLANDSKTCEDIDECIDILGSCSOHCYNMRGSRFCSCDTGYMLESDBRTCKVTASESLLLL 1439
QY 190 ---DS-T---A---RI-W---N---197
Db 1440 VASQNKIIADSVTSQVHNIYSLVENSYIIVAVDFDSISGRIFWSDATQGTWSAFQNGTD 1499
QY 198 ---LSB---N---S---T---203
Db 1500 RRVFPDSSIILTETIAIDWVGRNLYWTDYALETIEVSKIDGSHRTVLISKNLTNPRGLAL 1559
QY 204 ---S-G---P-T---207
Db 1560 DPRNEHLLFWSDGHHPRIERASMDGSMRTVIVQKIFWPCGLTIDYPNRLLYFMDSYL 1619
QY 208 ---Q---LVLRH---213
Db 1620 DYMFCDYNGHRRQVIASDLIIRHPVALTLFEDSVYTTDRATRVVRANKWHGQSVV 1679
QY 214 ---CI-R---EG---G---219
Db 1680 MYNIOMPLGIVAVHPKQPNVNCAPSRCSHLCLLSQGFPHFYCVCPGWSLSPDLLN 1739
QY 220 ---Q---D-VP-S-N-K-D---227
Db 1740 CLRDDQFLITVQHIIFGISLNPVKNSDAMVPIAGIQNGLDVEFDABEQIYIYVENPG 1799
QY 228 ---V---T---LDW---N---S-E---G---T 238
Db 1800 EIHVRKTDGTRNVFASISVMGFSMLALDWISERNLYSTNPTQSIÉVLTLHGDIYRKT 1859
QY 239 LLA-T---G-S---Y---DG---P---248
Db 1860 LIANDGTALGVGFPIGITVDPARGLYWSQGTDSGVPAKIASANMDGTSVKTLFTGNLE 1919
QY 249 ---A---R-I---W---T---253
Db 1920 HLECVTLDBEQKLYWAVTGRVIERGNVDGTRMILVHOLSHPWGIAVHDSFLYYTDEQ 1979
QY 254 ---K-D-G-N-L---ASTL-G-Q-H---K---266
Db 1980 YEVIERVDTGANKIVLRDNVNLRLGLQVYHRRNAESNGCSNNMACQIICLPVPGG 2039
QY 267 ---G---P---I-PALK---273
Db 2040 LPSCACATGKLPNDRSCSPYNSFIWVMSLAIRGFSLELSDHSETMVPVAGQGRNALH 2099
QY 274 ---W---N-K---K---GN- 279
Db 2100 VDVVSSGFIYWCDFSSVASDNAIRRIKPDGSSLMNIVTHGIGENVRGIAVDWVAGNL 2159
QY 280 ---F---281
Db 2160 YFTNAFVSETLIEVLINTYRRVLLKVTVMPRHIVVDPKRYLFWADYQRPKIERSF 2219
QY 282 ---L-S---A---G---VD---287
Db 2220 LDCTNRTVLVSEGIYTPRGLAVDRSDGYVYVWDDSLDIARIINGENSEVIRYGRYPT 2279

QY 288 ---K---T-T-I-I-W---D---294
Db 2280 PYGITYFENSIIWDRNLKKIIFOASKEPENTTEPTVIRDNINWLRDVTIFDKQVQPRSPA 2339
QY 295 ---A---HT---G---E---A---300
Db 2340 EVNNPCLENNGCGSHLCFALPGLHTPKDCAPGTLQSDGKNCAISTENFLIFALSLSLR 2399
QY 301 ---K---Q---Q-F---P- 305
Db 2400 SLHLDPENHSPPTTINVERTVMSLDYSVSDRIYFTQNLAGVGOISYATLSGIHTPT 2459
QY 306 ---F---306
Db 2460 VIASGIGTAGIADFWITRIIYSDYLNQMINSMAGDSNRTVIARVPKPRAIVLDPGCG 2519
QY 307 ---H---S-AP-ALD---313
Db 2520 YLYWADWDTHAKIERATLGNFRPVIWNSLVMPGLTLDYEDLLYVWDASLQRIERST 2579
QY 314 ---VD---W---Q---S---N---319
Db 2580 LTGVDRVIVNAVAHAFGLTYQYIYTDLYTQRIYRANKYDGSQGIAMTTLNLSQPRG 2639
QY 320 -NT---F---322
Db 2640 INTVVKQKQOCNPNCEQFNGGCSHICAPGNCAECQCPHEGNWYLANNRKHCIVDNGER 2699
QY 323 -AS---CS---TD-M-C-IH---V---333
Db 2700 CGASSPTCSNGRCISBEWKDNDNDGDSDEMESVCALHTCSPTAFTCANGRCVOYSYR 2759
QY 334 CKL---G-OD---R-P-I---342
Db 2760 CDYNDGDSDEAGCLFRDCNATTFMCMNRRCI PREFICNGVDNCHDNTSDEKNCPD 2819
QY 343 ---K---T---FQ---346
Db 2820 RYCQSGYTKCHNSNICIPRYLDCGNDGNSDENPTYCTHTCSSEFQACASGRCIPO 2879
QY 347 ---GH-T---349
Db 2880 HWYCDQETDCFDASDEPASCCHSERTCCLADEFKDCGRCIPSEWICDGDNDGMSDEDK 2939
QY 350 ---N-E-VNA---IK-W---357
Db 2940 RHQCNQNCSDSEFLCVNDRPPDRRCIPOSWVCDGVDCTGDYDENQNCNTRTTCSENEFT 2999
QY 358 ---357
Db 3000 CGYGLCIPKIFRCDRHNDGSDYSDERGCLYQTCQOQFTCQNGRCISKTFVCDEDNDCGD 3059
QY 358 -D---PT---G---N-L---363
Db 3060 GSDLMHLCHTPEPTCPPEFKCDNGRCIEMMKCNHLDCLDCLNSDEKKGGINCHEDPSI 3119
QY 364 ---LAS-CS---D-D-M---371
Db 3120 SGCDHNCTDLTSTFYCSCRPGYKLMDSKRTCDIDECTEMPFCVCKCENVIKCA 3179
QY 372 ---T---L---374
Db 3180 PGYLREPKTCRQNSNIEPYLIFSNRYLNLITIDGYFSLILEGLDNVVALDFORVEK 3239
QY 375 ---I---W---S---377
Db 3240 RLYWIDTQRIEMFLNKTNETIINHRLPAESLAVDWVSRKLYWLDARLDGLFVSDDL 3299
QY 378 ---M-KQ-D-N-C---VHDLO---A-H---NK---392
Db 3300 NGHREMLQACHCVDANNTFCFONPRGLALHP-QYGYLYWADWGHRAIYIGRVGMDGNTKSV 3358

QY 393 -----E-----I-YT-----IK----- 398
Db 3359 IISYKLEWPNGITDYTDNLLYWAHLYIEYSDLEGHHRHTYVDGALPHPPFAITIFED 3418
QY 399 -----W-----S-P-T--G-- 403
Db 3419 TIYWDNTRTEKGNKYDGSNRQTLVNTTHRPFDIHVYHPYQPIVSNPGTNGGCSH 3478
QY 404 -----PG-----T-----NN----- 408
Db 3479 LCLIKPGGKFTCEPDDFRTLQLSGSYCMPCMSSTQFLCANNEKCIPIWKKDQKDC 3538
QY 409 -----P-----NA-N-----LML-----AS- 417
Db 3539 SDGDELALCPQRCRLGQFCSDGNCSTSPOTLCNAHQNCPDGSDERL-LCENHHCDSN 3597
QY 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-----DV- 429
Db 3598 EWQCANRKCIPESWQCOTFNDCEDESDSHCASRTRCPQGFRCANGRCIPQAWKCDVD 3657
QY 430 -DRG-----I-CI--H-----T-L----- 438
Db 3658 NDCGDSDEPTEECSSAHLCDNFTEPSCNTYRCIPKAWCVNGVDDCRDSDGQCEER 3717
QY 439 T-----K-H-----Q-----E--P-----V----- 445
Db 3718 TCHPVGDFRCKNHHCIPLRWQCDQNDGDNDEENCAPRECTSEPRCVNQOCIPSRWI 3777
QY 446 -----Y-----S--V-----GS-----F----- 462
Db 3778 CDHYNDCGDSNDRDCERMTCHPEYFQCTSHCVHSELKCDGSDCLDASDEADCPTRF- 3836
QY 452 PDGRYL-AS----- 462
Db 3837 PDGAYCATWFECKNHVCIPPYWKCDGDDCGDGBELHLCLDVPNCNPRFRCDNNRC 3896
QY 463 -----D-----K-----CV-H--I-----W 469
Db 3897 IYSHEVCNVDGDDGDTDEEHCKRPTPKPCTEYKCGNGHCIPHDNVCDADDGCDW 3956
QY 470 -----N-----T-----Q--T-----G-----A-----L-V----- 477
Db 3957 SDELGCNKGKERTCAENICQNCQTLNEGGFICSTAGFETNVFDRTSCLDINECEQFGT 4016
QY 478 -----H-----SY----- 481
Db 4017 CPQHCRTNKGSYECVACDGTSMDSRDKRCRCAEGSSPLLLLPDNRVIRKYNLSSEFSE 4076
QY 482 -----G-----T-----G-----I--FE-----V-C-- 490
Db 4077 YLQDEEYIOAVDWDPKDGLSVVYTVRGEGRFAGAKRAVTPNFESGRNNLVQEVDL 4136
QY 491 -----W-----N-----AA----- 494
Db 4137 KLYVMQPDGIADVWGRHYIWSVKNRIEVAKL DGRYRKWLITDLDQPAIAVNPXL 4196
QY 495 -----D-----K--V----- 498
Db 4197 GLMFWTDWGKEPKIESAMNGEDBNILVFDLGNWPTGLSIDYLNDRYWSDFKEDVIETI 4256
QY 499 -----G-----A-SA-----S-----D-----G----- 505
Db 4257 KYDGTDRRVIAKEAMNPFYSLIDEPDQLYWISKEKEVWQKNGKQKKEKTLVVPNLQ 4316
QY 506 -----SV--C--V--L-----D----- 512
Db 4317 VRIFHQLRYKSVFNLCKQICSHLCLLRPGYSCACPDQSSFTGSGTTECDAAIPLINL 4376
QY 513 -----R-----K 514
Db 4377 PPRCRMHGNGCYFDETDLPK 4397

RESULT 5

US-08-652-877-84
; Sequence 84, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Rask, Lars
; APPLICANT: Crumley, Gregg R.
; APPLICANT: Morse, Clarence C.
; APPLICANT: Murray, Edward M.
; APPLICANT: Hjaln, Goran
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
; TITLE OF INVENTION: Thereof and DNA Encoding Same
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd., 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426-0107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.5.1
; SOFTWARE: Word 6.0 (Patentin)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,877
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: PCT/US95/15203
; APPLICATION NUMBER: PCT/US95/15203
; FILING DATE: 22-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,836
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,314
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitzky, Martin
; REGISTRATION NUMBER: 29,699
; REFERENCE/DOCKET NUMBER: A1355E-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-454-3816
; TELEFAX: 610-454-3808
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4655 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-652-877-84

Query Match 70.3%; Score 2577.8; DB 3; Length 4655;
Best Local Similarity 9.8%; Pred. No. 4e-32; Indels 3892; Gaps 381;
Matches 431; Conservative 59; Mismatches 20;

QY 1 M-----S-----I-----S--SDE- 7
Db 1 MDRGPAAVACTLLALVACLAPASGQECDSAFHRCGSGHCI PADWRCDGTGKDCSDDAEI 60
QY 8 ---VN---F-----L---V----- 12
Db 61 GCAVTCQGYPKQSEGGCIPSSWVCDQDQDDGSDERQDCSQSTCSSHQTCSNGQC 120
QY 13 ---YR-----Y-----L-----Q-----E---S 19
Db 121 IFSEYRCDHVRCDPDGADENDCOYTCBQLTCDNGACYNYSKCDWKVDCRSDSDEINCT 180
QY 20 -----G-FS-----H-----S---A--F-T-----F---G--I----- 30

Db 181 EICHNEFCGNGECIPRAYVVDHNDQDGSDEHACNYPTCGGYQFTCPSGRCIYQNWV 240
QY 31 -----S-----H-----IS-----Q-SNI 39
Db 241 CDGEDCKUNGDEGGESGPHVHKSPREWSPEGRGCIYKVCDDGILDCPREDEN - 299
QY 40 N---G-----AL---VP-----PAALJ---S-----I-I-- 53
Db 300 NTSTGKYCSMTLCSALNCQYQCHETPVGGACFPGPYIINHNSRVCVEFDCCQINGICD 359
QY 54 QK-----G-L--OY-----VE 61
Db 360 QKCESRGRHLCHCEBYILRGYCKANDSPGEASIFNSGRDLLIGDIHGRSPRIIVE 419
QY 62 -----A-----E---VSI-----N 67
Db 420 SQNRGVAVGFHYHLQRVFTVQNKVFSVDINGLNIQEVLSVETPENLAVDWVN 479
QY 68 -----E-----DG---TL----- 72
Db 480 KIVLVETKVNRIQVNLGDSYRVTLITENLGHPRGIAVDPTVGYLFFSDWESLSGPKLE 539
QY 73 --P-DG--R-----P-----IE-----S 80
Db 540 RAFMDGSRKDLVKTKLGNPAGVTLDMISKRVYWDVSRFDYIETVTDGIRQKTVVHGG 599
QY 81 L-----SL---I--D--AVM----- 88
Db 600 LIHPHFGVSLFEGQVFTDTRKAVLKANKFTETNPQVYQASLRPYGVTVYHSLRQPYA 659
QY 89 --P-D--V-V-----Q--T----- 94
Db 660 TNPCKDNNGGCEQCVLSHRTDNDGLGFRCKCTFGQLDTERHCAVQNFLIFSSQVAI 719
QY 95 R-----Q----- 97
Db 720 RGIPFTLSTQEDVMVPSGNPFPVGDIDFDAQDSTIFFSDMSKHMIFKQKIDGTGREILA 779
QY 98 -----A-----Y--R-DK--LAQ-----Q--H-A-- 108
Db 780 ANRVENVESLAFDWISKNLWTDSHYKSIWMLADKTRTVQVYLNPNRVSUVHPFAGY 839
QY 109 -----A-----A-----AA-----A----- 114
Db 840 LPFTDWRPAKIRAWSGSHLLPVNTTLGWENGLAIDWAASRLYWDAYDPKIEHSTP 899
QY 115 -----A-----A-----A-----T----- 118
Db 900 DGLDRRLGHIEQWTHPFGLAIFGEHLFTTDWRLGAIIRVRKADGGMETVIRSGIAYILH 959
QY 119 -----N-OQGS-A-----KNG-----EN-----TA-N-----G----- 133
Db 960 LKSYDNIQTGSNACNQTHPNPGDCHSHCFPPVFNFORVCGCPYGMRLASNHLTCBGDPTN 1019
QY 134 E---E-----N-----GA--H-----T----- 140
Db 1020 EPTTEQGLFSFPCKNGRCVPNYLDCGDVDDCHNSDEQLCGTLNNTCSSAFTCGHGEC 1079
QY 141 I-A-----N-----H-----T-DMME-----V-D--GD-- 154
Db 1080 IPAHWRCKDKENDCVGSDENHCPHAPASCLDTQYTCDNHQICISKWVCDTNDNDCGDGSD 1139
QY 155 -----VEI--PS-----N-----K-A----- 162
Db 1140 EKNCNSTETQPSQFQNCNPHRCIDLFSVCDGDKDCVDGSDVGVCLNCTASQPKCASGDK 1199
QY 163 -----V-----VL-----R-G-H-----E----- 169
Db 1200 CIGVTNRCDGVFDCSDNSDEAGCTPPGNGCHSDEFCQEDGICIPNFWECDHGPDCLYG 1259
QY 170 S-E-----V-----F-----IC--AW-----N-----P----- 179
Db 1260 SDEHNACVPKTCPSYFHCNNGNCIHRAMLCDRDNDCGMSDEKDCPTQPFRCPSWQOC 1319

QY 180 -----V-----SD----- 182
Db 1320 LGHNCVNLVVCDDGIFDCPNGTDESPLCNGNSCDFNGGCTHECVQEPGAKCLCPG 1379
QY 183 LL-----V-----S-----GS-----G----- 189
Db 1380 LLANDSKTCEDIDECIDILGSCSOHCYNMRGSRFCSCDTGYMLESDDGRTCKVTASESILLL 1439
QY 190 -----DS-T-----A-----RI-W-----N-- 197
Db 1440 VASQNKIADSVTSQVHNIYSLVNGSYIVAVDFDSISGRIFWSDATOGKTWAFQMGTD 1499
QY 198 -----LSE-----N-----S-T----- 203
Db 1500 RRVVPDSSIIETETIAIDVGNRLYWDYALETIEVSKIDGSHRTVLISKNLINPRGLAL 1559
QY 204 -----S-G-----P-T----- 207
Db 1560 DPRMNEHLLFWSDGWHHPRIERASMDGSMRTVIVQDKIFWPCCGLTIDYPNRLLYFMDSYL 1619
QY 208 -----Q-----LVLH----- 213
Db 1620 DYMFCDYNGHRRQVIASDLIIRHPYALTLPEDSVYWTDRTRVRMRANKWHGGNQSVV 1679
QY 214 -----CI--R-----EG-----G----- 219
Db 1680 MYNIQWPLGIVAVHPKQNSVNPFCASRCSHLCLLSSQGPHEFVSCVPSGWSLSPDLLN 1739
QY 220 -----Q-----D--VP-S--N--K-D----- 227
Db 1740 CLRDDQPLITVRQHIIFGISLNPVKSDAMVPIAGIQNGLDVEFDABQYIYVVENPG 1799
QY 228 -----V-----T-----S--LDW-----N--S-E--G-----T 238
Db 1800 EHRVKTGTNRVTFASISMVGPSMNALDMSRNLSTYNPTQTSLEVLTLHGDIYRKT 1859
QY 239 LLA--T-----G-S-----Y-----DG-----F----- 248
Db 1860 LIANDGTALGVGPPIGITVDPPARKLYWSDQGTDSGVPAKIASANMDGTSTVKLTG 1919
QY 249 -----A-----R-I-----W-----T----- 253
Db 1920 HLECVTLIDEEQKLYWAVTGRVIERGNVDGTDMLVHQLSHPWGIAVHDSFLYTYDEQ 1979
QY 254 ---K-D--G-N--L--ASTL-G-Q--H-----K----- 266
Db 1980 YEVIKVDKATGANKIVLRDNVNLRLQVYHRRNAESNGCSNNMNACQICLPVPGG 2039
QY 267 -----G-----P-----I--FALK----- 273
Db 2040 LFSACATGFKLNPDRNSCPYNSPIVVSMLSAIRGFSLESDHSETWVPVAGQGRNALH 2099
QY 274 -----W-----N-----K-----K-----GN- 279
Db 2100 VDVVSSGYIWCDFSSVASDRAIRRIKPDGSSLMNIVTHGICENGVRGIAVDWVAGNL 2159
QY 280 -----F-----I----- 281
Db 2160 YFTNAFVSTLIEVLRINTYRRVLKVTVDMPRHIVDPKRYLFWADYGQRPKIERSF 2219
QY 282 -----L-S-----A-----G--VD----- 287
Db 2220 LDCNTNVLVSEGIVTTPRGLAVDRSDGYVYVVDVDSLDIARIRNGENSEVIRGSRPT 2279
QY 288 -----K-----T-----T-I--I-W-----D----- 294
Db 2280 PYGITVFENSIWVDRNLKKIIOASKEPENTPTVIRDNINWLDRDVTIFDKQVQPRSPA 2339
QY 295 -----A-----HT-----G-----E-----A----- 300
Db 2340 EVNNPCLNNGGCSHLCPALPGLHTPKCDCAFGLTQSDGKNCAISTENFLIFALSNSLR 2399

301 QY -K- -Q- -Q- -F- -P- 305
2400 Db SLHLDPNHSPFOTINVERTVMSLDYSDVSDRIYFTQNLASGVQISYATLSSGHTPT 2459
306 QY -F- -F- -F- -F- -F- 306
2460 Db VIASGIGTAGIADWITRRIYSDYLNQMSMAEDGSNRTVIARVPKRAIVLDPDQCG 2519
307 QY -H- -H- -H- -H- -H- 313
2520 Db YLYWADWDTHAKIERATLGNFRVPIVNSLVMPGSLTLYEEDLLYWDASLQRIERST 2579
314 QY -VD- -VD- -VD- -VD- -VD- 319
2580 Db LTGVREVINAAVHAFGLTYGYIYTDLTYRIYRANKYDGSQGIAMTLLSOPRG 2639
320 QY -NT- -NT- -NT- -NT- -NT- 322
2640 Db INTVVKQKQCNPCQFNGGCSHCAPGPGABECQCPHEGNWYLANNRKHCIVDNGER 2699
323 QY -AS- -AS- -AS- -AS- -AS- 333
2700 Db CGASSFTCSNGRCISEEWKCDNDNDGDSDEMESVCAHTCSPTAFTCANGRCVOYSYR 2759
334 QY -CKL- -CKL- -CKL- -CKL- -CKL- 342
2760 Db CDYNDGDSGDSAGCLFRONATTEPMCNRRCIIPREFICNGVDNCHDNTSDKNCPD 2819
343 QY -K- -K- -K- -K- -K- 346
2820 Db RTCQSGYTKCHNSNICIPRVVLCDGNDGNDSDENPTYCTHTTCSSEFQCSAGRCIPQ 2879
347 QY -GH- -GH- -GH- -GH- -GH- 349
2880 Db HWYCDQETDCEPASGCSHRTCLADEFKDGRGRCIPSEWICDGDNDGDSDEDK 2939
350 QY -N- -N- -N- -N- -N- 357
2940 Db RHQONQCSDSFELCVNDRPPRCIPQSVWGDVDTGDYDENQONCTRTCSNEFT 2999
358 QY -D- -D- -D- -D- -D- 357
3000 Db CGYGLCIPKIFCDRHNDGDSYDERGCLYQTCQONQFTCQNGRCISKTEVCEDNDNDG 3059
358 QY -D- -D- -D- -D- -D- 363
3060 Db GSELMHLCHTPTCPPEHFKDNGRCIEBWKLCNHLDCLDNSDEKSGCINECHDPSI 3119
364 QY -LAS- -LAS- -LAS- -LAS- -LAS- 371
3120 Db SGCDHNCITLTSFYCSRGYKMSDKRTCVDIDECTEMPFVCSQKCNVIGSYCKCA 3179
372 QY -T- -T- -T- -T- -T- 374
3180 Db PGVLRPDGKTCRQNSNIEPYLFSNRYRLNLTIDGYFYSILEGLDNVALLDFRVEK 3239
375 QY -I- -I- -I- -I- -I- 377
3240 Db RLYWIDTQROVIERMFLNKTKETIINHLRPAESLAVDWVSRLYWLARDLGLFVSDL 3299
378 QY -M- -M- -M- -M- -M- 392
3300 Db NGGHRMLAQCHVDANNFTCFDNPRLALAP-QYGLYWADWGHRAVIGRVGMDGNTKSV 3358
393 QY -E- -E- -E- -E- -E- 398
3359 Db IISTKLEWPNGITIDYNDLLYWADAHGLYIEYSDLEGGHRTVYDGLPFPFAITIFED 3418
399 QY -W- -W- -W- -W- -W- 403
3419 Db TIYTDWNTVTEKGNKYDGSNRQTLVNTTHRPFDIHVHPYQPIVSNPCGNNCGCSH 3478
404 QY -PG- -PG- -PG- -PG- -PG- 408

3479 Db LCLIKPGGKFTCECPDDFRTLQLSGSTYCMPCMSSTQFLCANNEKCIPIWKKDQKDC- 3538
409 QY -P- -P- -P- -P- -P- 417
3539 Db SDGSDALALCPQFRLGQFCSDGNCSTPQTLNAHQNCPCDGSDEDL- -LCENHHCDN 3597
418 QY -A- -A- -A- -A- -A- 429
3598 Db EQCANKRCIPESWQCDTENDCEDNDESDSHCASRTCRPGQFCRANGRCIPOAWKCDVD 3657
430 QY -DRG- -DRG- -DRG- -DRG- -DRG- 438
3658 Db NDCGDSHDEPIECMSAHLCDNFTBFSCKTNRICPKAVCVGVDDCRNSDEQCEER 3717
439 QY -K- -K- -K- -K- -K- 445
3718 Db TCHPVGDFRCKNHHCIPLRWQCDGQNDGDSDEENECAPRECTESBFCVNCQICPSRWI 3777
446 QY -Y- -Y- -Y- -Y- -Y- 451
3778 Db CDHYNDGDSNDRDCMRCTCHPEYFQCTSGHCVHSELKCDGSDACLDASDEADCPTRF- 3836
452 QY -PDGRL- -PDGRL- -PDGRL- -PDGRL- -PDGRL- 462
3837 Db PDGAYCOATWFECKNHVCIPPYWKCDGDDDDGDSDEELHLCLDVPNSNPRFCDNNRC 3896
463 QY -D- -D- -D- -D- -D- 469
3897 Db IYSHEVNGVDDCGDGTDEBHCRTKPTKPCTEYEVKCGNGHCCI PHDNVCDADDGCDW 3956
470 QY -N- -N- -N- -N- -N- 477
3957 Db SDELGNKGRKTCBAENICEQNTQLNEGFCISCTAGFETNVFDRTSCLDINECEQFGT 4016
478 QY -H- -H- -H- -H- -H- 481
4017 Db CPOHCNTKSYECVACDGTMSDRPGKRCABGSSPLLLPDNVIRKYNLSSEFSE 4076
482 QY -T- -T- -T- -T- -T- 490
4077 Db YLQDEBYIOAVDYDWDPKDIGLSVYTVRGEGRFGAKRAYIPNFESGRNNLVQEVDL 4136
491 QY -W- -W- -W- -W- -W- 494
4137 Db KLYVMQPGIADVWGRHIYNSDVKNKRIEVAKLDGRYRKWLITDLPQAAIAVNPKL 4196
495 QY -G- -G- -G- -G- -G- 498
4197 Db GLMFWTDWKEPKIESAWNGEDRNILVFEGLGWPTGLSIDYLNNDRIYWSDFKEDVIET 4256
499 QY -A- -A- -A- -A- -A- 505
4257 Db IKYDGTDRRVIKAEAMNPYSLDIFEDQLYWSKEGEVWQKNGKQKKEKTLVNPWLT 4316
506 QY -SV- -SV- -SV- -SV- -SV- 511
4317 Db QVRIFHQLRYNKSVPNLCKQICSHLCLRLRPGGYSCACPOGSSFIEGSTTECDAAIELPIN 4376
512 QY -R- -R- -R- -R- -R- 514
4377 Db LPPPCRCMHGNGCYFDETDLPK 4398

RESULT 6

US-08-652-877-88
; Sequence 88, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Rask, Lars
; APPLICANT: Crumley, Gregg R.

QY 190 ---DS-T---A---RI-W---N---197
Db 1440 VASOKIIADSVSOVHNIYSLVNGSYIIVAVDFDSISGRIFWSDATOGKTWAFQNGTD 1499
QY 198 ---LSE---N---S-T---203
Db 1500 RRVVFDSSIIILTETIAIDWGRNLWTDYALETIEVSKIDGSHRTVLISKNLTPRGAL 1559
QY 204 ---S-G---P-T---207
Db 1560 DPRNHELLFWSMGHHPRIERASMDGSMRTIVQDKIFWPCGLTIDYPNRLLYFMDSYL 1619
QY 208 ---Q---LVLRH---213
Db 1620 DYMDPCDYNHRRQVIAISDLIIRHPYALTLFEDSVYVTDRAVRVRANKWGHGNSVV 1679
QY 214 ---CI-R---EG---G---219
Db 1680 MYNIQWPLGIVAVHPKQPNVNPAPSRCSHLCLSSQGFHFVSCVCPGWSLSPDLLN 1739
QY 220 ---Q---D---VP-S---N---K-D---227
Db 1740 CLRDDQFLITVRQHIIFGILNPEVKSNDAMPVIAIGIQLDVEFDDAEQYIYVENPG 1799
QY 228 ---V---T---S---LDW---N---S-E---G---T 238
Db 1800 EIHVRKTDGNTIRTVFASISVMGPMALADWISNLSTNPTQSEVLTLHGDIYRKT 1859
QY 239 LLA---T---G-S---Y---DG---F---248
Db 1860 LIANDGTALGVGPIGITVDPARGKLYWSQDGTSGVPKAKIASANMDGTSVKTLFTGNLE 1919
QY 249 ---A---R-I---W---T---253
Db 1920 HLECVTLDIEBQKLYWAVTGRVIERGNVDGTRMILVHQLSHPWGIAVHDSFLYTDEQ 1979
QY 254 ---K-D---G-N---L---ASTL-G-Q-H---K---266
Db 1980 YEVIERVDXATGANKIVLRDNVFNRLGLQVYHRRNAEBSNGCSNNMACQOICLPVPGG 2039
QY 267 ---G---P---I---PALK---273
Db 2040 LFSCACATGKLPNDNRSCSPYNSFIVVMSLAIRGFSLESDHSETMVPVAGQGRNALH 2099
QY 274 ---W---N---K---K---GN- 279
Db 2100 VDVVSSGFIYWCDFSSVASDRAIRRIKPDGSSLMNIVTHGENGVRGIAVDWVAGNL 2159
QY 280 ---I---F---I---281
Db 2160 YFTNAFVSETLIEVLRINTTYRRVLLKVTVDMPRHIVVDPKNRYLFWADYQRPKIERSF 2219
QY 282 ---L-S---A---G---VD---287
Db 2220 LDCTNRTVLVSEGIIVPRGLAVDRSDGVYVWVDDSLDIARIINGENSEVIRYGSRYPT 2279
QY 288 ---K---T---T---I---I-W---D---294
Db 2280 PYGITVFENSIIWDRNLKFIQASKEPENTEPPTVIRDINWLDRVTIFDKQVQPRSPA 2339
QY 295 ---A---HT---G---E---A---300
Db 2340 EVANNPCLENNGCSHLCPALPGLHPTKDCARGTQSDGKNCAISTENFLIFALSLSLR 2399
QY 301 ---K---Q---F---P---305
Db 2400 SLHLDPENHPPTQINVERTVMSLDYSDVSDRIYFTQNLASGVQISYATLSSGIHTPT 2459
QY 306 ---F---306
Db 2460 VIASGIGTAGIADWITRIIYSDYLNQMINMAEDGSNRTVIARVPKRAIVLDPCCQ 2519
QY 307 ---H---S---AP---ALD---313

Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSSLVMPSGITLDYEEDLLYVWDASLQRIERST 2579
QY 314 ---VD---W---Q---S---N---319
Db 2580 LTGVDEVIIVAAVHAFGLTLYQYIYVTDLYTORIYRANKYDGSQGIAMTTLLSQPRG 2639
QY 320 ---NT---F---322
Db 2640 INTVKNQKQCCNPNCEQFNGGCSHICAPGPNABEQCEHGNWYLANNRKHCIIVNGER 2699
QY 323 ---AS---CS---TD-M---C-IH---V---333
Db 2700 CGASFTCSNGRCISEWKCDNDNDGSDGSESVICALHTCSPTAFTCANGRCVQYSYR 2759
QY 334 CKL---G---QD---R-P-I---342
Db 2760 CDYNDGCGSDEAGCLFRDCNATTEFCMNRRCIPREFICNGVDNCHDNNTSDEKNCPD 2819
QY 343 ---K---T---T---PQ---346
Db 2820 RTCQSGYTKCHNSNICIPRYLDCGNDGCDNSDENPTYCTHTCSSEFQCSGRCIPO 2879
QY 347 ---GH---T---349
Db 2880 HWYCDQETDCFADSPASCGHSERTCLADEFKDGGRCIPSEWICDNDNDGCDMSDEK 2939
QY 350 ---N---E---VNA---IK-W---357
Db 2940 RHOCQONCSDESFLCVDNDRPPDRRCIPOSVWVGVDCTDGDYDENQNCNCTRTTCSNEFT 2999
QY 358 ---LAS---CS---D-D-M---371
Db 3120 SGCDHNCTDLTSFYCSCRPGYKMSDKRKTCDIDECTEMPFVCSQKCNENIGVSIKCA 3179
QY 372 ---T---L---374
Db 3180 PGYLREPDGKTCRQNSNIEPYLIFSRYLRLNLTIDGYFYSLLILEGLDNVVALDFRVEK 3239
QY 375 ---I---W---S---377
Db 3240 RLYWIDTQRVIERMFLNKTNETIINHRLPAESLAVDWVSRKLYWLDARLDGLFVSDL 3299
QY 378 ---M-KQ---D-N-C---VHDLO---A---H---NK---392
Db 3300 NGHREMLAQHCVDANNTFCFDPNRLALHP-QYGYLYWADWGHAYIGRVGMDGTNKS 3358
QY 393 ---E---I-VT---IK---398
Db 3359 IISTKLEWNGITIDYNDLLYWADAHLYEYSDLEGGHRRHTVYDGLALPHPAITIFED 3418
QY 399 ---W---S-P-T---G---403
Db 3419 TIYWTDMNTRTVEKGNKYDGSNRQTLVNTTHRPFDIHVYHPYRQPIVNSPCGTNNGCSH 3478
QY 404 ---PG---T---NN---408
Db 3479 LCLIKFEGGKFTCECPDPFRTLQLSGSTYCMPCMSSTQFLCANNKRCIPIWKKCDGQKOC 3538
QY 409 ---P---NA-N---LML---AS- 417
Db 3539 SUGSDEALCPORFCKLGFQCSGDNCTSPQTLCAHQNCPOGSDERL-LCENHHCDSN 3597
QY 418 ---A---S---F---DS---T---VR---L---W---DV- 429

Db 3598 EWQANKRCIPESWQDFTFNDCEDESDSHCASRTCRPGQFRCANGRCIPOAWKCDVD 3657
Qy 430 -DRG-----I--CI--H-----T-L-----438
Db 3658 NDCGDHSDPIECMSAHLCDNFTEBSCKTNKPIKWAICVGVDDCRDSDQCEER 3717
Qy 439 T-----K--H-----Q-----E--P-----V-----445
Db 3718 TCHPVGDFRCKNHHCILPRWQCGDNDGDSDEBENCAPRECTESBRCVNCQICPSRWI 3777
Qy 446 -----Y--S--V-----A-----FS 451
Db 3778 CDHYNDGDSRDERDCMRTCHPEYFOCTSGHCVHSELKCDGSADCLDASDEADCPTRF- 3836
Qy 452 PDGRYL-AS-----GS-----F-----462
Db 3837 PDGAYCAQATFECKNHVCIPYKWKCGDDCGDSDEELHCLDVPNCSPNRCDNNRC 3896
Qy 463 -----D-----K-----CV-H--I-----W 469
Db 3897 IYSHVCGVDDCGDGTDETEEHCKRTPKPTCTEYKCGNGHCIPHNDVNCDDADDGCDW 3956
Qy 470 -----N-----T-----Q--T-----G-----A-----L-V-----477
Db 3957 SDELGNKKGKERTCAENICQNCQTLNNEGGFICSTAGFTNVFDRTSCLDINECQFGT 4016
Qy 478 ---H-----SY-----T-----G-----I--FE-----V-C---490
Db 4017 CPQHCRTKGSYECVACADGFTMSDRPKCAEGSSPLLLLPDNRVIRKYNLSERFSE 4076
Qy 482 -----G-----T-----G-----I--FE-----V-C---490
Db 4077 YLQDEEYIAVDYDWDPEDIGLSVWYTVRGEGRFGAIAKRAVIPNFESGRNNLVQEVLD 4136
Qy 491 -----W-----N-----AA-----494
Db 4137 KLYVMQPDGIADVWGRHYNSDVKNKRIEVAKLGRYKWLISTDLQPAIAVNPXL 4196
Qy 495 -----G-----G-----D-K--V---498
Db 4197 GLMFWTDWGKPKESANWNGEDBNILVFDLQWPTGLSIDYLNDRYWSDFKEDVIET 4256
Qy 499 ---G-----A-SA---S---D-----G-----505
Db 4257 IKYDGTDRRVIAKEAMPYSLDIFEDQLYWISKEGEWKNQKFGQKKEKTLVVPWLT 4316
Qy 506 -----SV---C-V---L-----D-----511
Db 4317 QVRIFHQLRYNKSVPNLKQICSHLCILLRPGYSCACPOGSSFIEGTTTECDAAIELPIN 4376
Qy 512 L-----R-----K 514
Db 4377 LPPPCRMHGGNCYFDETDLPK 4398

RESULT 7

US-08-652-877-90
; Sequence 90, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Rask, Lars
; APPLICANT: Crumley, Gregg R.
; APPLICANT: Morse, Clarence C.
; APPLICANT: Murray, Edward M.
; APPLICANT: Hjalms, Goran
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
; TITLE OF INVENTION: Thereof and DNA Encoding Same
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd., 3C43

; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426-0107
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.5.1
; SOFTWARE: Word 6.0 (Patentin)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,877
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/15203
; FILING DATE: 22-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,836
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,314
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitzky, Martin
; REGISTRATION NUMBER: 29,699
; REFERENCE/DOCKET NUMBER: A1355E-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-454-3816
; TELEFAX: 610-454-3808
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4655 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-652-877-90

Query Match 70.3%; Score 2577.8; DB 3; Length 4655;
Best Local Similarity 9.8%; Pred. No. 4e-32;
Matches 431; Conservative 59; Mismatches 20; Indels 3892; Gaps 381;
Qy 1 M-----S-----I-----S-----S--SDE- 7
Db 1 MDGPAAVACTLLALVALCLAPASGOECDSAHFRCSGHCIPADWRCDGTGKDCDDADEI 60
Qy 8 ---VN---F-----L---V-----12
Db 61 GCAVTCQGYFKQSEGGCIPSSWVYCDQDQDCDDGSDERQDCSQTCSHQITCSNGQC 120
Qy 13 ---VR-----Y---L-----Q-----E---S 19
Db 121 IPSEVRCDHVRCDPGADENDCQYPTCEQLTCDNGACYNYSQKCDWKVCDRSDSDINCT 180
Qy 20 -----G-FS-----H-----S---A--F-T-----F---G-I-----30
Db 181 EICLHNEFSCNGECIPRAYVCDHNDQDGSDEHACNYPTCGGYQFTCPSGRCIYQNWV 240
Qy 31 -----ES-----H-----IS-----Q--SNI 39
Db 241 CDGEDDCKNGDEDCESGPHDVHKSPREWSCPESGRCISYKVCVDGILDGCFREDEN- 299
Qy 40 N--G-----AL-----V-----PAALI-----S-----I--I-- 53
Db 300 NTSTGYKSMTLCSALNCOYQCHETPYGGACFCPPGYIINHNSRVCVFEDDCQINGICD 359
Qy 54 QK-----G--L---QY-----VE 61
Db 360 QKCESRPGRLHCHCEBEGYILERQYCKANDSFGASIIIFSNGRDLLIGDIHGSRFLVE 419
Qy 62 -----A-----E---VSI-----N 67
Db 420 SQNRGVAVGVAFYHLQRFVFTDVTQNVKFSVDINGLNQIEVLNVSFETPENLAVDWNV 479

QY 68 ---E---DG---TL--- 72
Db 480 KIVLVETKVRIDWNLDGSRVTLITENLCHPRGIAVDTVGYLFFSWESLSGEPKLE 539
QY 73 --F-DG--R---P---IE---S 80
Db 540 RAFMDGSRNRDLVKTKLGFAGVTLDMI SKRVYWDVSRFDYIETVTVYDGIQRKTVVHGS 599
QY 81 L---SL---I---D---AVM--- 88
Db 600 LIHPFGVSLFEGQVFTDWTMAVLKANKFTETNPQVYYQASLRPYGVTVYHSLRQPYA 659
QY 89 --P--D---V-V---Q--T--- 94
Db 660 TNECKNNGGCEQVULSHRTDNDGLGRCKTFGFGQDTHDRHCTAVQNFILFSSQVAI 719
QY 95 R-----Q----- 97
Db 720 RGIPFTLTQEDVMVPVSGNPFVGDIDFAQDSTIFFSDMSKHMIFKQIDGTGREILA 779
QY 98 ---A---Y---R--DK---LAO---Q---H--A-- 108
Db 780 ANRVENVESLAFDWISKNLYWTDSHYKXISVMSRLADKTRTVVQYLYNPNRSVVVHPFAGY 839
QY 109 ---A---A---A---AA---A--- 114
Db 840 LPFTDWRPAPAKIMRAMSDGSHLLPVINTLWGPNGLAIDWAASRLYVWDAYFDKIEHSTF 899
QY 115 ---A---A---A---A---T--- 118
Db 900 DGLDRRLGHIEQWHPFLAIFGEHLFTDWRLGAIIRVKADGEMTVIRSGIAVILH 959
QY 119 ---N--QGS-A---KNG---EN---TA-N---G--- 133
Db 960 LKSYDNIQTGSAACNOPTHENGDCSHFCFPVFNQVRCVCGPYGMRLASNHLTCBGDPTN 1019
QY 134 E--E---N---GA--H---T--- 140
Db 1020 EPTEQCGLSFPCKNGRCVNPYYLGDVDDCHNSDEQLCGTLNNTCSSAFTCGHGE 1079
QY 141 I-A-----N---H---T--DMME-----V-D---GD--- 154
Db 1080 IPAHWRCKRNDKVDGSDHNCPTHAPASCLDTQYTCDNHQCSIKWVCDTNDNCGDSD 1139
QY 155 ---VEI--PS---N---K-A--- 162
Db 1140 EKNCNSTETCQSFQNCNPHRCIDLSEVCDGDKDVCVDSDEVGVCLNCTASQPKASGDK 1199
QY 163 ---V---VL---R--G--H---E--- 169
Db 1200 CIGVTNRCDGVFDCSDNSDEAGCFTRPPGCHSDEFQCEGICIPNFWCEDGHPDCLYG 1259
QY 170 S-E---V---F---IC--AW---N---P--- 179
Db 1260 SDEHNACVPKTPSSYFHCNDGNCIHRAWLCDRNDGDMDSDEKDCPTQPFRCPSQWQC 1319
QY 180 -----V-----SD----- 182
Db 1320 LGHNICVNLVVDGIFDCPNGTDESPLCNGNSCDFNGGCTHECVQEPFGAKCLPLGF 1379
QY 183 LL-----V--S---GS----- 189
Db 1380 LLANDSKTCEDIDECIDLGSQHCYNMRGSRFCSDTYGMLSDGRTCKVTASESLLLL 1439
QY 190 ---DS-T---A---RI-W---N--- 197
Db 1440 VASQNKIADSVTSQVHNIYSLVENGSYIYAVDPDSISGRIFWSDATQGTWAFQNGTD 1499
QY 198 ---LSE---N---S---T--- 203
Db 1500 RRVFDSIIILTETIADWGRNLYWTDYALETIEVSKIDGSHRTVLISKMLNPRGLAL 1559
QY 204 -----S--G-----P---T--- 207

Db 1560 DPMNEHLLFWSDGWHHPRIERASMDGSMRTVIVQDKIFWPCGLTIDYENRLLYFMDSYL 1619
QY 208 -----Q-----LVLRH----- 213
Db 1620 DYMFCDYNGHRRQVIASDLIIRHPYALTLPEDSVYWTDRATRRVMRANKMHGNGOSVV 1679
QY 214 -----CI--R---EG-----G----- 219
Db 1680 MNYIQWPLGIVAVHPKQPNVNPFCAFSCSHLLSSQPHFYSCVCSGWSLSPDLLN 1739
QY 220 -----Q-----D--VP-S--N---K--D--- 227
Db 1740 CLRDDQPLITVRQHIIFGILNPEVKSNDAMVPIAGIQNGLDVEPDAAEQIYVWENPG 1799
QY 228 --V---T---S---LDW---N---S-E---G---T 238
Db 1800 EHRVKTGDTNRTVFASISMVGPSMLALDWISRLYSTNPTQSTEVLTGLHGDIRYKT 1859
QY 239 LLA--T---G-S---Y-----DG---F--- 248
Db 1860 LIANDGTALGVGPGIITVDPARGKLYWSDQGTDSGVPAKIASANMDGTSVKTLFTGNLE 1919
QY 249 -----A-----R-I---W-----T--- 253
Db 1920 HLECVTDIDIEQKLYWAVTGRGVIERGNVDGTDRLVHQLSHPMGIAVHDSFLYTTDEQ 1979
QY 254 --K--D--G-N--L--ASTL-G-O--H---K--- 266
Db 1980 YEVIERVKATGANKIVLRDNPVNLRLGLOVYHRRNAESNGCSNNMNAQQICLPVPGG 2039
QY 267 -----G-----P-----I--FALK----- 273
Db 2040 LFSCACATGFKLPNDRSCSPNSPIVSMLSAIRGFSLELSDHSETMVPVACQGRNALH 2099
QY 274 -----W-----N---K-----K-----GN- 279
Db 2100 VDVVSSGFYWCDFSSVASDNAIRRIKPDGSSLMNIVTHGIGENGVRGIAVDWVAGNL 2159
QY 280 -----F-----I----- 281
Db 2160 YFTNAFVSETLIEVLINTYRRVLLKVTVDMPRHIVVDPKRYLFWADYQRPKIERSP 2219
QY 282 --L-S-----A--G---VD----- 287
Db 2220 LDCTNETVLSGIVTPRGIAVDRSDGYVYVWVWDSLDIARIINGENSEVIRVGSRYPT 2279
QY 288 -----K---T---T-I--I-W---D----- 294
Db 2280 PYGITVFENSIIWDRNLKKIFOASKEPENTEPPTVIRDININWLRDVTIPDKQVQPRSPA 2339
QY 295 -----A---HT-----G-----E---A--- 300
Db 2340 EVNNPCLENNGCCHLCPALPGLHTPKDCARGTLQSDGKNCAISTENFLIFALSNSLR 2399
QY 301 -----K---Q-----O--F-----P- 305
Db 2400 SLHLDPENHSPPTQINVERTVMSLDYSDSDRIYFTQNLASGVGOISYATLSSGIHTPT 2459
QY 306 -----F----- 306
Db 2460 VIASGIGTADGIAFDWITRRIYVSDYLNQMINSMAGDSNRTVIARVPKPRAILVDPCCQ 2519
QY 307 -----H-----S--AP---ALD--- 313
Db 2520 YLWAWDWDHAKIERTATLGNFRVPIVNSLSVWPSGLTDYEEDLLYWDASLQRIERST 2579
QY 314 ---VD---W---Q-----S---N--- 319
Db 2580 LTGVREVIVNAAVHAFGLTYGYIWTDLTYRIYRANKYDGSQGIAMTLLSOPRG 2639
QY 320 -NT-----F----- 322

Db 2640 INTVVKQKQCNPCQFNGCGSHICAPGFGAECQCPHGWNLYANNRKHICIVDNGER 2699
Qy 323 --AS--CS-----TD-M--C-IH-----V----- 333
Db 2700 CGASSTCSNGRCISBEWKCDNDGDSDEMESVCAHTCSPTAFTCANGRCVQYSYR 2759
Qy 334 CKL-----G--QD-----R-P--I----- 342
Db 2760 CDYDNGGDSDEAGLFRDCNATTFMCMNRRCIPREFICNGVDNCHDNTSDEKNCPD 2819
Qy 343 -----K-----T-----PQ----- 346
Db 2820 RTCQGYTKCHNSNICIPRYLDCGNDGDNSENPTYCTTHTCSSEFQCSAGRCIPQ 2879
Qy 347 -----GH--T----- 349
Db 2880 HWYCDQETDCFDASBPASCGHSERTCLADEFKDGGRCIPSEWICGDNDCGMSDEK 2939
Qy 350 -----N--E--VNA-----IK-W----- 357
Db 2940 RHQONQNCSDSEFLCVDNRPDRRCIPQSWVCDGVDCTGDYDENQNCNTRTCSNEFT 2999
Qy 358 ----- 357
Db 3000 CGYGLCIPKIFCRDRHNDGYSDBERGCLYQTCQNOFTQNGRCISKTFVCDENDCGD 3059
Qy 358 --D-----PT-----G-----N-----L----- 363
Db 3060 GSDLMHLCHTPEPTCPPEHFKDNGRCIEMMKLCHLDCLDNDSEBKGGCINECHDPSI 3119
Qy 364 -----LAS--CS-----D-D--M----- 371
Db 3120 SGCDHNCTDLTFSFYCSRPGYKLMKDKRKTVDIDECTEMPFVCSQKENCNIGSYICKCA 3179
Qy 372 -----T-----L----- 374
Db 3180 PGYLREPDGKTQNSNIEPYLFSNRYLRNLTIDGYFYSLEGLDNVVALDFRVEK 3239
Qy 375 -----I----- 377
Db 3240 RLYWIDTQVIERMFLNKETIINHRLPAESLAVDWVSRKLYLDARLDGLFVSDL 3299
Qy 378 -----M-KQ--D--N--C-----VHDLO--A--H-----NK-- 392
Db 3300 NGHRRMLAQHCVDANNTFCFNDPRGLALHP-QVGYLYADWCHRAVIGRVGMDGNTKSV 3358
Qy 393 -----E-----I-YT-----IK----- 398
Db 3359 IISTKLEWPNGITIDYTNLLYWAHAHLGYIEYSDLEGHHRHTVYDGLPFPFAITIFED 3418
Qy 399 -----W----- 403
Db 3419 TIYWTWNTRTVEKNGKNDGSRQTLNTHTRFPDIHVTHPYQPIVSNPCGTNGGCSH 3478
Qy 404 -----PG-----T-----NN----- 408
Db 3479 LCILKFGGKFTCECPDFTLQSGSTYCMPCWCSSTQFLCANNEKCIPIWKKCDQKDC 3538
Qy 409 -----P-----NA--N-----LML-----AS-- 417
Db 3539 SDGSDALCPQRFRLGQFCSDGNGCTSPOTLCNAHQNCPCDGSDEDL-L-CENHHDCSN 3597
Qy 418 -----A-----S-----P-----DS-----T-----VR-----L--W--DV-- 429
Db 3598 EWOCAKRCIPESWQCDTNDCEDNDEDESSHCASRTCPGQRCANGRCIPQAWKCDVD 3657
Qy 430 -DRG-----I--CI--H-----T-L----- 438
Db 3658 NDCGDSDEPIEBECMSAHLCONFTEFSCKTNRYCIPKAVCNVGDVDDCRDSDQCEER 3717
Qy 439 T-----K--H-----Q-----E--P-----V----- 445
Db 3718 TCHPVGDFRCKNHHCIPLRWQCDGQNDGDNSENCAPRECTSEBFCVNOQCIPSRWI 3777

Qy 446 -----Y-----S--V-----A-----FS 451
Db 3778 CDHYNDGNSDSDCEMTCHEPYFQCTSGHCVHSELKCGSADCLDASDEADCTRF- 3836
Qy 452 PDGRYL-AS-----GS-----F----- 462
Db 3837 PDGAYCOATMFECKNHVCIPPYWKCDGDDCGDGSDEELHCLDVPFNSPFRCDNNRC 3896
Qy 463 -----D-----K-----CV-H--I-----W 469
Db 3897 IYSHEVCNGVDDCGDGTDETEHCRKPTPKPCTEYKCGNGHCIPHDNVCDADDCGDW 3956
Qy 470 -----N-----T-----Q--T-----G-----A-----L-V----- 477
Db 3957 SDELGCNKERTCAENICEQNCTQLNEGFFICSTAGFETNVDFRTSCLDINECBQFGT 4016
Qy 478 --H-----SY-----R----- 481
Db 4017 CPQHCRNTRKGSYECVACDGTSMDSRPGKRCAGSSPLLLLPDNRVIRKYNLSRFRSE 4076
Qy 482 -----G-----T-----G-----I--FE-----V-C-- 490
Db 4077 YLQDEBYIOAVDYDMDPEDIGLSVVYTVRGSGRFGAIKRAYIPNFESGRNNLVQEVDL 4136
Qy 491 -----W-----N-----AA----- 494
Db 4137 KUKYMQPDGIAVDMVGRHIYMSDVKNKRIEVAKLDRYRKWLISDLDQPAIAVNPKL 4196
Qy 495 -----G-----D-K--V----- 498
Db 4197 GLMFWTDMGKEPKLESAMNGEDRNILVFPEDLGWPTGLSIDYLNNDRIYMSDFKEDVIET 4256
Qy 499 -----G-----A-SA--S-----D-----G----- 505
Db 4257 IKYDGTDRVIAKEAWNPSLDIFEDQLYWSKEGVEWQKNGKQKKEKTLVNPWL 4316
Qy 506 -----SV-----C--V-----L-----D----- 511
Db 4317 QVRIHQRLYKNSVNPMLCKQICSHLCLLRPGGYSCACPOGSSFIEGTTTECDAAIELPIN 4376
Qy 512 L-----R-----K 514
Db 4377 LPPPCRCMHGNCYFDETDLPK 4398

RESULT 8

US-08-652-877-86

; Sequence 86, Application US/08652877

; Patent No. 6187548

; GENERAL INFORMATION:

; APPLICANT: Akerstrom, Goran

; APPLICANT: Juhlin, Claes

; APPLICANT: Raek, Lars

; APPLICANT: Crumley, Gregg R.

; APPLICANT: Morse, Clarence C.

; APPLICANT: Murray, Edward M.

; APPLICANT: Hjalms, Goran

; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments

; TITLE OF INVENTION: Thereof and DNA Encoding Same

; NUMBER OF SEQUENCES: 106

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Rhone-Poulenc Rorer Inc.

; STREET: 500 Arcola Rd., 3C43

; CITY: Collegeville

; STATE: PA

; COUNTRY: USA

; ZIP: 19426-0107

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Macintosh

; OPERATING SYSTEM: System 7.5.1

; SOFTWARE: Word 6.0 (patentin)

Db 1680 MYNIQWPLGIVAHPSKQPNVNCAPSRCSHILCLSSQGHFYFVSCVCPSGWSLSPDLLN 1739
Qy 220 -----Q-----D--VP-S--N--K-D-----227
Db 1740 CLRDDQPLITVRQHIIIFGILSNPEKSNDAWPIAGIQNGLDVEPDDAEQYIYWENPG 1799
Qy 228 -----V-----T-----S-----LDW-----N-----S-E-----G-----T 238
Db 1800 EHRVKTGTNRVTFASISMGVSPMNLALDWISRLNSTYNTPTQSIIEVLTLHGDIYRKT 1859
Qy 239 LLA---T-----G-S-----Y-----DG-----F-----248
Db 1860 LIANDGTALGVGPPIGITVDPARKLYWSDQGTSDGVPAKIASANMDGTSVKTLFTGNLE 1919
Qy 249 -----A-----R-I-----W-----T-----253
Db 1920 HLECVTLDBEQKLYWAVTGRGVIERNVDGTRMILVHQLSHPWGIAVHDSFLYYTDEQ 1979
Qy 254 ---K-D---G-N---L---ASTL-G-Q-H-----K-----266
Db 1980 YEVIERVKATGANKIVLRDNVNLRLGLQYHRRNAEASSNGCSNNNAQCQICLPVPGG 2039
Qy 267 -----G-----P-----I---FALK-----273
Db 2040 LFSCACATGKLPNDNRSCSPYNSFIVVMSLAIKRGFSLSDHSETWVPVAGQGRNALH 2099
Qy 274 -----W-----N-----K-----K-----GN- 279
Db 2100 VDVVSSGFIYWCDFSSVASDNAIRIKPDGSLMNIIVTHGIGENGVRGIAVDWAGNL 2159
Qy 280 -----I-----F-----I-----281
Db 2160 YFTNAFVSETLIEVLINTTYRRVLLKVTVDMPRHIVVDPKNRYLFWADYQRPKIERSF 2219
Qy 282 -----L-S-----A-----G-----VD-----287
Db 2220 LDCTNRTVLVSEGIPTPRGLAVDRSDGYVYVWDSLDIIAIRINGENSEVIRYGRYPT 2279
Qy 288 -----K-----T-----T-I---I-W-----D-----294
Db 2280 PYGITVPENSIWDRNLKXIFOASKEPENTPEPTVIRDNINWLROVTFDKOVQPRSPA 2339
Qy 295 -----A-----HT-----G-----E-----A-----300
Db 2340 EVNNNPCLENGGSHLCPALPGLHTPKCDCAFGTLQSDGKNCAISTENFLIFALSNSLR 2399
Qy 301 -----K-----Q-----Q-----F-----P-----305
Db 2400 SLHLDPENHSPPQTINVERTVMSLDYSVSDRIYFTQNLASGVGQISYATLSSGIHTPT 2459
Qy 306 -----F-----306
Db 2460 VIASGIGTAGIADWITRIIYSDYLNQMINSMEDGNSRNTVIARVPKPRAILVDPCCQ 2519
Qy 307 -----H-----S-AP-ALD-----313
Db 2520 YLYWADWDHAKIERATLGNFRVPIVNSLVPSGLTLDYEEDLLYWWDASLQRIERST 2579
Qy 314 ---VD-----W-----Q-----S-----N-----319
Db 2580 LTGVREVI VNAAVHAFGLTYGYIYWTDLTYQRIYRANKYDGSGLAMTTNLLSQPRG 2639
Qy 320 -NT-----F-----322
Db 2640 INTVVRNQKQCNPCQFNGGSHICAPGNGAECQCPHEGNWYLANNRKHICVIDNGER 2699
Qy 323 ---AS---CS-----TD-M---C-IH-----V-----333
Db 2700 CGASSFTCSNGRCISEBWKCDNDNDGSDGDSMBSCVLCALHTCSPTAFTCANGRCVQYSYR 2759
Qy 334 CKL-----G---QD-----R-P---I-----342
Db 2760 CDYNDGSDGSDRAGCLFRDCNATTEPMCNRRRCIPREFICNGVDNCHDNTSDEKNCPD 2819

Qy 343 -----K-----T-----FQ-----346
Db 2820 RTCQSGYTKCHNSNICIPRYVLCDGNDCGDNSDENPTYCTTHTCSSSEFQCTSGRCIPQ 2879
Qy 347 -----GH---T-----349
Db 2880 HWYCDQETCFDASDEPASCGRHSERTCLADBFKCGGRCIPSEWICDGDNDGDMGSDDEK 2939
Qy 350 -----N---E---VNA-----IK-W-----357
Db 2940 RHOCQONQSDSEFLCVNDRPPRRRCIPQSWVWCDGVDCDGDYDENQNCNTRTCSENEFT 2999
Qy 358 -----357
Db 3000 CGYGLCIPKIFRCDRHDCGSDYSDRGCLYQTCQOQFTCQNGRCISKTFVCDENDCGD 3059
Qy 358 ---D-----PT-----G-----N-----L-----363
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLDNDSEKGCINECHDPSI 3119
Qy 364 -----LAS---CS-----D-D---M-----371
Db 3120 SGCDHCTDTLTSFYCSCRPGYKMSDKRXTCDIDECTEMPFVCSOKCENVIYSICKCA 3179
Qy 372 -----T-----L-----W-----K 374
Db 3180 PGYLRPDGKTCRONSNIPEYLIFSNRYLRNLITIDGYFYSLLILEGLDNNVALDFORVEK 3239
Qy 375 -----I-----W-----S-----377
Db 3240 RLYWIDTQVIERMELFNKTNKTIINHRLPAAESLAVDWVSRKLYWLADRLDGLFVSDL 3299
Qy 378 -----M-KQ---D-N---C-----VHDLO-----A---H-----NK---392
Db 3300 NGHREMLAQHCVDANNTCFDNPRGLAHP-QYGYLYWADWGHRAIYIGRVMDGTNKS 3358
Qy 393 -----E---I-YT-----IK-----398
Db 3359 IISTKLEWNGITIDYTNDLLYWADAHGLGYIEVSDLEGHHRHTYDGLPHPFAITIFED 3418
Qy 399 -----W-----S-P-T---G---403
Db 3419 TIYTDWNTRTVEKNGKYDGSNRQTLVNTTHRPDIHVHYPRQPIVSNPCGTNNGCSH 3478
Qy 404 ---PG-----T-----NN-----408
Db 3479 LCLIKPGGKFTCECPDPFRTLQLSGTYCMPMCSSTQFLCANNEKCIPIWVKDQKOC 3538
Qy 409 -----P---NA---N-----LML-----AS- 417
Db 3539 SDGSDELALCPQBFCLGQFCSDGNCSTPOTLCNAHQNCPOGSDERDL-LCENHHCDN 3597
Qy 418 ---A-----S-----F-----DS-----T-----VR-----L---W---DV- 429
Db 3598 EQCANKRCIPESWQCDTENDCEDNDESDSHCASRTCPQFCRANGRCIPQAWKCDVD 3657
Qy 430 -DRG-----I---CI---H-----T-L-----438
Db 3658 NDCGDHSDPEIBECMSSAHLCDNFTEFCKTNRRCIPKWAVCNGVDDCDRNSDEQGCER 3717
Qy 439 T-----K-H-----Q-----E---P-----V-----445
Db 3718 TCHPVGDFRCNHHCIPLRWQCGQNDGNDSEENCAPECTESBFRVCVNOQCIPSRWI 3777
Qy 446 -----Y---S---V-----A-----FS 451
Db 3778 CDHYNDGNSDERDCEMTCHPEYQCTSGHCVHSELKCDGSADCLDASDEADCTRF- 3836
Qy 452 PDGRYL-AS-----GS-----P-----462
Db 3837 PDGAYQATWFECKNHVCIPIYWKCDGDDCGDGSDEELHLCLDVPNCNRRFCNDNRC 3896

QY 463 -----D-----K-----CV-H-I-----W 469
Db 3897 IYSHEVNGVDDCGDGTDETEHCRKTPPKCTEYKYKNGCHIPHNDVNCDDADDGCDW 3956
QY 470 -----N-----T-----Q--T-----G-----A-----L-V----- 477
Db 3957 SDELGNCKGERICAENICQNCQCTQNLNEGFGICCTAGFETNVFDRTSCLDINECQFGT 4016
QY 478 -----H-----SY-----R----- 481
Db 4017 CPQCRNTKGSYECVACADGFTSMRPGKRCACAAEGSSPLLLLPDNRKIRKYNLSSEFSE 4076
QY 482 -----G-----T-----G-----G-----I--PE-----V-C--- 490
Db 4077 YLQDEEYQAVDWDPDXDGLSVVYTVRGEGRFGAIKRAIYIPNFESGRNNLVQEVDL 4136
QY 491 -----W-----N-----AA----- 494
Db 4137 KLYVMQPDGIADVWGRHIYNSDVKNKRIEVAKLDGRYKWLJSTDLQPAALAVNPKL 4196
QY 495 -----G-----D-K--V--- 498
Db 4197 GLMFWDGKPEKXESAMNGEDRNILVFDLWPTGLSIDYLNDRYNSDFKEDVIET 4256
QY 499 -----G-----A-SA-----S-----D-----G----- 505
Db 4257 IKYDGTDRRVIKAMNPYSLDIFEDQLYWISKEGEVWQKNGFGQKKEKTLVNPWLT 4316
QY 506 -----SV-----C--V-----L-----D----- 511
Db 4317 QVRIFHQLRYNKSVNPLCKQICSHLCLLRPGGYSCACPGSSFIEGSTECDAAIELPIN 4376
QY 512 L-----R-----K 514
Db 4377 LPPPCRMHGNCYFDETDLPK 4398

RESULT 9

US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING I9G Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-388-9
Query Match 69.8%; Score 2560.4; DB 3; Length 5405;
Best Local Similarity 9.6%; Pred. No. 2e-31;
Matches 410; Conservative 81; Mismatches 17; Indels 3766; Gaps 353;
QY 1 M-----S--I-----SS-D-EVN-----FLV-----YR-----Y----- 15
Db 1 MGALWSWILWAGATLLGLTQEASVDLK-NTGREBFLTAFLQNYLAYSKAYPRLLISS 59
QY 16 -----L-Q-----ES-----G--F--S--H----- 23
Db 60 LSESPASVILSQADNTSKVTVRPGESVVMNISAKAEMIGSKIFOHAVVIHSDYALSIVQ 119
QY 24 -----SA-----F-----T 27
Db 120 ALNAKPDPTAELTLRLPIQALGTEYFVLTPPGTSARNVKFAVVAGAAGASVSLKGSVT 179
QY 28 F-----G-----I 30
Db 180 FNGKFYPAGDVLRTVLQPNVAQLQSSVDLSGSKVTASSPVAVLSHSCAQKHTTCNHVV 239
QY 31 E-----S--H-I-----SQS-----N--I----- 39
Db 240 EQLLPTSAWGHYVVPFTLASQSRDYDLAFVVASOATKLTYNHGGITSGRGLQAGDVVEFEV 299
QY 40 -----N-----GAL--V--P--P--AAL-----I--S----- 51
Db 300 RPSWPLYLSANVGIQVLLFGTGAIKNEVTVDPVLVLPDVAACPAVYVVKSVPGCEGVAL 359
QY 52 II-Q-K--G-----IQY--V--E--AEV--S--I-----N----- 67
Db 360 VVAQTKAISGLTIDGHAVGAKLTWEAVPGSEFSYAELGTADMIHTABATNLGLITFG 419
QY 68 --E-----D-G-T-L----- 72
Db 420 LAKAIGYATAADCGRTVLSFVPEPSCEGMCQAQRCQVVGKGAGCVAESTAVCRAQGDPH 479
QY 73 ---FDGR-----P--IE----- 79
Db 480 YTTFDGRRYDMGTCSTYTWVLCSEDDTLPAFSVEAKNEHRSRRVSVYGLVTVRAYSHS 539
QY 80 -SLS-----LID--A--V-----M-----P--D-----VW-----Q----- 93
Db 540 VSLTRGEVGFVLVDNQRSLPVSLSRGLRVYQSGPRAVVELVFLVTVTYDWDCLALS 599
QY 94 -TR-Q-QA-----Y-----RD-----KL-----A 104
Db 600 PARFQDVCGLCGNYNGDPADFLTPDGLAPDAVEFASWKLDDGDLCEGCGQNNCPA 659
QY 105 ---Q-QH-----A----- 108
Db 660 CTGQAQHYEGDRLCGMLTKLDGPFVAVCHDTLDPFLEQCVYDLCVVGGERLSLGRGLS 719
QY 109 A-A-A-----A-----A-----AAA-TN----- 119
Db 720 AVAQACLELGISVGDWRSPANCPLSCPANSRYELCGPACPTSCNGAAAPSNCSGRPCVEG 779
QY 120 -----Q-Q--G-S--A-K-----NGEN--T----- 130
Db 780 CVCLPGFVAGGACVPASCGCTFGQLQAPGEVWADLCCQRRCTCNGATHQVTRCDKQ 839
QY 131 ---ANGE---EN-----GA--H-----T-----I-----AN- 143
Db 840 SCPA-GERCSVQNGLLGCYDPDRFGTCQSGSDPHYVVSFGDGRFDFMGTCVTLLVSGSCQNA 898
QY 144 -----N-H-----T-----DMME----- 150
Db 899 ALPAFRVLVENEHRGSGTYSYTRAVRVEARGVKVAVRREYPGQVLDVDDVLQYLPFOAADG 958

QY 151 -----V--D-G--D--V--EIPS----- 159
Db 959 QVQVFRQGRDAVVRTDFGLAVTYDMNARVTAKEPSSVAEALCGLGNFNGDPADDLALRG 1018
QY 160 -----K----- 161
Db 1019 GGOAANALAFGNWQEBTRPGCGATEPGDKLDSLVAQLOQSKNECGILADPKPPREC 1078
QY 162 -----AV---VL-R-----G--H----- 168
Db 1079 HSKLDPOGAVRDCVYDECLLPQSGPLCDALATYAACQAAGATVHPWRSEELCPISCPP 1138
QY 169 -----ES----- 170
Db 1139 HSHYEACSYCLPLSCGDLPPVGGCGSECHGCVCEGDFALSGESCLPLASCGCVHQGVTH 1198
QY 171 -----E-----V----- 172
Db 1199 PPGQTFYPGCGDSLCHQCBGLVSCSSCGPHEACQPSGSLGCVAGSSTCOASGDP 1258
QY 173 -----F-----I-----C-----AW----- 177
Db 1259 HYTFDGRDFMGTCVYVLAQTCGTRPGLHRFAVLQENVANGNRVSVTRVITQVANF 1318
QY 178 -----N-----PV-----SD-----L-----L-----VS- 186
Db 1319 TLRLEQRMKVTVNGVDMKLPVVLANGQIRASQHSQSDVVIETDFGLRVAYDLVYVRVTV 1378
QY 187 -----G--SGD-----S--TARI-----W----- 196
Db 1379 PGNYQOMGLGNGYNDPKDFQKPNQSOAGNA-NEFGNSWEEVVPDSFCLPPTPCPPG 1437
QY 197 -----N----- 197
Db 1438 SEDCIPSHKCPPELEKYYQKEEFCGLLSSTGTPSLSCHKLVDPPGPKDCIFDLCLGGN 1497
QY 198 LS-----E--NS-----T----- 203
Db 1498 LSLCSNIHAYVSACQAAGHVEPWRTETPCMECPNPNHVELCADCATCSLGCALSAPQ 1557
QY 204 -----SG-----PTQ-----L-----VL-----R----- 212
Db 1558 QDGCBAEGCQDSGLYNGOACVPIQQCGCYHNGVYEPETVLIDNCRQOCTCHAGKM 1617
QY 213 -----H-----C-----I-----R-----E-GGO----- 220
Db 1618 VCOEHSCKPGQVQCPQSGGILSCVTDKPCHGVTCRPOETCKEQQGGQGVCLPNEYATCWLWG 1677
QY 221 -----D-----V-----P--S----- 224
Db 1678 DPHYHSFDGRKDFQGTCTNYVLATTCGPGVSTQGLTPFTVTTKQNRGNPAVSYYRVVTV 1737
QY 225 -----N-----KD-----VT-----SLDWN- 234
Db 1738 AALGTNISHKDEIGKVRVNGVLTALPVSVADGRISVTOGASKALLVADFGQLQVSYDWNW 1797
QY 235 -----S----- 236
Db 1798 RVDVTLPSHYGAVGCLGMDRNPNDQVFPNGTLAPSIPIWGSWRAPGWDPLCWDEC 1857
QY 237 -G--T-----L-----LA--TG----- 243
Db 1858 RGSCTCPEDRLBQYEGPGFGLAPGTGGPFTTCHAHVPPESFFKPGKCVLDVCMGGGDRD 1917
QY 244 -----SY----- 245
Db 1918 ILCKALASYAACQAAGVIEDWRAQVGCETCPENSHYEVCGPPCPASCPSAPLITTEA 1977
QY 246 -----D-GF--A-R-I-----W---T-----K---DGN----- 257
Db 1978 VCBGPVCEGQCDAGFVLSADRCPLNNGCGCWANGTYHAGSEFWADGTCSQWCRCPG 2037
QY 258 ---L---AS-----TL--GOHKG--PIF-ALK--W-----N----- 275

Db 2038 GGSIVCTPASCGLGEVCGLLPSGQH-GCQPVSTA-ECQAWGDHYPVTLTLDHRENFQGTCE 2095
QY 276 -----KK-G--NF-----I-----LSA-----G 285
Db 2096 YLLSAPCHGPPLGAENFTVTVANEHRGSAVSYTRVTLQIYNHSLTLSARWPKLQVDG 2155
QY 286 V-----DK-----TT----- 290
Db 2156 VFVTLFPQLDSLHLHLSGADVVTITTSGLSLAFDGDSPVLRVPAAYAGSLGCLGNYN 2215
QY 291 ----- 292
Db 2216 QDPADDLKAVGKPAQWQVGGAGCGECVSKPCPCTPEQBSFGPGDACGVISATDGP 2275
QY 293 -----W---D----- 294
Db 2276 LAPCHGLVPPAQYFQGCCLLDACQVQHGGCLCPAVATYAAQQAQURRWRPDCPPF 2335
QY 295 ---AHT-----G-E-A-K-----Q----- 302
Db 2336 QCPAHSHELCDGSDCPSGLSAPESACREBGCVCADGFFVLSGDTCVPVCGCCLHD 2395
QY 303 -----Q-F----- 304
Db 2396 DRYYPGLQTFYFPGPGCDLCLCRCEGGEVSCPESSCGPHETCRPSGSLGCVAGSTTQCA 2455
QY 305 ---P-----F-----H-----S----- 308
Db 2456 SGDPHYTFDGRDFMGTCVYVLAQTCGTRPGLHRFAVLQENVANGNRVSVTRVITVQ 2515
QY 309 -A-----P-----ALD----- 313
Db 2516 VANFTLRLEQRMKVTVNGVDMKLPVVLANGQIRASQHSQSDVVIETDFGLRVAYDLVYV 2575
QY 314 -V-----D-----W----- 316
Db 2576 RVTVPNGYQLMCGLCNGYNDPKDDFKPNGSQAGNANEFNGSWEEVVPDSCLPPTC 2635
QY 317 -----Q----- 317
Db 2636 PPSGECIPSEBCEPPELEKYYQKEEFCGLLSSTGTPSLSCHKLVDPPGPKDCIFDLCLG 2695
QY 318 -----SN-----N-TF-----A--S--CS--T- 327
Db 2696 GGNLSILCSNIHAYVSACQAAGHVEPWRTETPCMECPNPNHVELCADCATCSLGCALS 2755
QY 328 -----D-----MC-I-----H-----V-----C--- 334
Db 2756 PLQCPDCAEGCQDSGLYNGOACVPIQQCGCYHNGVYEPETVLIDNCRQOCTCHAG 2815
QY 335 KL-----GO-----D-----RP-----I----- 342
Db 2816 KYVVCQHSCKPGQVQCPQSGGILSCVTDKPCHGVTCRPOETCKEQQGGQGVCLPNEYATCW 2875
QY 343 -----K-----T-F-----Q----- 346
Db 2876 LMGDPHYHSFDGRKDFQGTCTNYVLATTCGPGVSTQGLTPFTVTTKQNRGNPAVSYYRV 2935
QY 347 -----G-----HTNE-----VN-----A--I-----K----- 356
Db 2936 VTVAAALGTNISHKDEIGKVRVNGVLTALPVSVADGRISVAQASKALLVADFGQLQVSYD 2995
QY 357 --W--D--PT-----GNL-----LA-----S----- 366
Db 2996 MNWRVDTLPSSHYGAVGCLGMDRNPNDQVFPNGTLAPSIPIWGSWRAPGWDPLCW 3055
QY 367 --C--S-----D----- 371
Db 3056 DECRGSCPTCPEDRLBQYEGPGFGLAPGTGGPFTTCHAHVPPESFFKPGKCVLDVCMGGG 3115
QY 372 ---TL-K-----I--W-----SM----- 378

Db	599	RSWLLHTECLPEICAGPHLPHSTVSYSSENYSGWNRITIPLPNAALFTRNIRWQTGPI	658
Qy	121	-----Q-----G-S-----A-----K-----	125
Db	659	LGNWAIIDNVYIGPSCILKFCGRCQCTRHGCKCDPGFSGPACEMASQTFFMFISESFGSS	718
Qy	126	-----N-----G-E-----N-----	129
Db	719	RLSSYHNFYSIRGAEVSGCGVLASGKALVFENKEGRRQLITSLDSSQSRLQFTLRIGS	778
Qy	130	-----T-----A-----NG-----	133
Db	779	KSVLSTCRAPDQPGEGVLLHYSYDNGITWKLLEHYSYLSYHEPRIISVELPGDAKQFIQ	838
Qy	134	-----E-----E-----N-----G-----AH-----	139
Db	839	FRWQPHYSSOREDDVAIDILIMTSVLNSISLDFTNLVEVTQSLGFLGNVQPYCGHDW	898
Qy	140	-----T-----I-A-----N-----NH-----	145
Db	899	TLCFTGDSKLASSMRYVETQSMOIGASYMIQFSLVMGCGQKYTPHMDNQVKLEVSTNHGL	958
Qy	146	-----	146
Db	959	TWHLVQBECLPMPSCOEFTSASIYHASEFTQWRRVIVLLPQKTWSSATFRWSQSYTA	1018
Qy	147	-----D-----M-----M-E-----V-D-----G	153
Db	1019	QDEWALDSIYIGQOCPMCSGHGSCDHGICRCDOGYOGTECHPEALPSTIMSDFENQNG	1078
Qy	154	-----D-----V-----EI-P-----S-----NKA-----V-V-L-----	165
Db	1079	WESDWQVEIGGEIVKPPQCGGVSSGSLYFSKAGKQLVSWDLTDSWDFVQFYIQIGG	1138
Qy	166	-----R-----G-----H-----E-----SEV-----FI-----C-----	175
Db	1139	ESASCNKPDREGEVLLQYSNNGIOWHLLAEYFSDFKPRFYLELPAAKTPTCTRFR	1198
Qy	176	-----AW-----NPV-----S-D-----L-L	184
Db	1199	WQVPVFGSEDYQWANDDIIILSEKQKQIIPVINPTLPQNFYEKPAFDYPMNQMSVWML	1258
Qy	185	-----V-----S-----G-S-GD-----S-TAR-----I-----	195
Db	1259	ANEGWKNETFCATPSAMIFGKSGDGRFAVT-RDLTKPGYVLOFKLNGCANQFSSTA	1317
Qy	196	-----WNL-----S-----E-NS-----T-----SG-----P	206
Db	1318	PVLLQYSHDAGMSWFLVKEGYPASAGKCGEGNSRELSEPTWYHTGDFEWTTRITIVIPR	1377
Qy	207	-----T-----Q-----L-VL-----R-----H-CI-----	215
Db	1378	SLASKTRFRWIOESSQKQVPPFGLDGVYISEPFCPSYCSGHGDCISGVCFCDLGYTAAQ	1437
Qy	216	-----R-EG-----G-Q-----D-----D-----VP	223
Db	1438	GTCVSNVPNENMPDRFEGKSLPLYKITGAQVGTGCGTLDGKSLYFNGPGKREARTVP	1497
Qy	224	-----SN-----K-D-----	227
Db	1498	LDTRNIRLVQFYIQIGSKTSGITCIKPRTRNEGLVQYSNDNGILMHLLELDFMSPLEP	1557
Qy	228	-----VTS-L-L-D-----W-----N-SE-----	236
Db	1558	QIISIDLPODAKTAPATFRWQPHQKHSAQWALDDVLIGMNDSSQTFQDFKDGSDILQ	1617
Qy	237	-----G-----T-L-----LA-----T-----G-----	243
Db	1618	ANVYRIQGGQVDDCLSDNALIFTENIGKPRYAETWDFHVSASTFLQFMSMGCSKPPFS	1677
Qy	244	-----S-----Y-----D-----G-----FA-----RI-----	251

Db	1678	NSHSVQLQYSLNNGKDWHLVTBECVPTTICLHYTHTESSYTSRFPQWKRITVYLP	STI	1733
Qy	252	W-----T-----		253
Db	1738	SPRTRPRWIQANYTVGADSWAIDNVVLASGPMWCSGRGICDAGRVCVDRGFGPYCPV	P	1797
Qy	254	KD-----GNL-----A-----S-T-L-----G-----		263
Db	1798	VPLPSILKDDFNGNLHPDLMPVEYGAERGNLNGETIKSGTSLIFKGEGLRMLISRD	LDCT	1857
Qy	264	Q-----Q-----HK-----G-----P-----I-F-----		270
Db	1858	NTMVVQSLRFIAKSPERSHSILLOFSISGGITWHLMDFEYFPQTTNULFINVPLPY	TA	1917
Qy	271	A-----LKW-----N-KK-----GN-----F-----		280
Db	1918	QTNATRFRL-WQPVNCKKEEIIWVDDFIIDGNVANNPVMLLDTDFGPREDNWFY	PGG	1976
Qy	281	I-L-----SA-----G-V-----D-----KTTII-----		292
Db	1977	NIGLYCPYSSKGAPEDSAMFVSNEVGEHSITTRDLNVNENTIIIOFEINVGST	DSSA	2036
Qy	293	W-----DAH-----T-----GEAKQ-----Q-----F-304		
Db	2037	DPVRLFEPSRDFGATWHLPLLPLCYHSSHSVSSLCSTEHPFSSTYAG-TMGQWR	REVVHFG	2095
Qy	305	PFH---SA-----PALD---VDM-----Q-S-N-----N-T-----FA	323	
Db	2096	KLHLCSVRFRWYQGFYPAGSQPVTAIDNVYIGPQCEMCNCGSCINGTKICDP	GY	2155
Qy	324	SC-ST-----D-----M-----C-I-----		331
Db	2156	GPTCKISTKNPDLKDDFEGQLBSRFLMLSGGKPSKCGILSSGNLFPENEDGLR	MLMT	2215
Qy	332	H-----V-----C-K-----L-G-----QD-----R-340		
Db	2216	RDLDLSHAREVQPFMRIGCKGKGVDPSPQVLLQYSLNGGLSWSLQELFSS	NNVGRY	2275
Qy	341	PTK---T---FO---GH---T---NEV---NAI-----K	356	
Db	2276	IALEIPLKARGSGTRLRWQPSENGHYFSPWIDQILIGN-LSGNTVLEDDFT	ILDSRK	2334
Qy	357	W-----D-P---TG-----N---LL---ASCS---D---369		
Db	2335	WLLHPGCTKMPGVSGTGDALVFIEKASTRYVYSTDVAVNEDSLQIDFAA	SCVSDCYA	2394
Qy	370	DM-----T---L-KI-----W-----S---377		
Db	2395	IELEYSVDLGLSHWHLVRLCLPTNVCSRYHLQILVSDTFNKWTRITLPLPY	TRSOAT	2454
Qy	378	M-----KQ-----DN-----CV---H-----D---386		
Db	2455	RFRWQAPDPKQOTWAIIDNVYIGDCIDCMCSGHGRICQNCVCDQBMGGY	CDDPETS	2514
Qy	387	L-----Q-----A-H-----NKE-393		
Db	2515	PTQLKONFNPAPSSQNLWTVNGGKLSVCVAVASGHALHPSGGCSRLLVTL	DLNLTNAEF	2574
Qy	394	I---Y---T-----IKW-----S-PTG-----403		
Db	2575	IQFYFMYGCLITPNRRNQVLLSEYVNGGITWNLLEIFDYQSKP-GFVN	ILLPPDAKE	2633
Qy	404	P---G---TN---N-----P-NA-----411		
Db	2634	IATFRFWQPRHGLDQNDWAIDNVILSGADQRTVMLDTFSSAPVQHERSPAD	AGPVG	2693
Qy	412	N---L-----ML-----A-----416		
Db	2694	RIAFDMEDKTSVNEHWHFDHDDCTVERFCDSPGVMLCGSHDGREVYATH	DLTTEGV	2753
Qy	417	S---A-----S-F-----D-----S-----422		
Db	2754	IMOPKISVGCKVSEKIAONOIHOVYSTDFGVSNVILVPOCLPADPKCSG	SVSOPSFVFT	2813

QY 423 ---T---VR---L---W-D---VD---RG---IC 434
Db 2814 KGWKRTYPLPESLVGNPFRFYQKYSMDQWADNFYLPFGCLNCRGHGDCLEQCIC 2873
QY 435 ---I---HTL-T---KHQ---E-P---V---Y---S- 447
Db 2874 DPGYSGPNCYLTHLTKERFDEEIKPDLWMSLEGSTCTEGGILAEATYFGST 2933
QY 448 ---V---A-F---S-P---DG--- 454
Db 2934 VRQAVTDLRLGAKFLQWGRIGSNNMTSCHRPKRKEGVLLDYSDGIGITWLLHEM 2993
QY 455 ---R-Y-L---A---SG--- 460
Db 2994 DYQKYSVRHDYILLPEBALNTTTLRWQPFVISNGIVSGVERAQWALDNILIGGAEI 3053
QY 461 ---S---FD---K---C---VHI-W-N---T--- 471
Db 3054 NPSQLVDFDDEGTSHENWFSFYPNAVRTAGFCGNPSPHLYWPNKKDKTHNALSRELI 3113
QY 472 ---Q---T-G---A---IV--- 477
Db 3114 IQGYMMQFKIIVGCEATSCGLHSMVLEXTKDARSWSQLVOTQCLPSSNSIGCSPPQ 3173
QY 478 -H-S-Y---R---G-T--- 483
Db 3174 FHEATYNSVSSWKRTIOLPDHVSSATQFRWIKQBEETEKQSWAIDHVIYIGEACPK 3233
QY 484 ---G---G-I---FE-VC---W-N--- 492
Db 3234 LCSGHGYCTTCAICIDESFGDDCSVPSHDLPSYIKDNFESARVTEANWETIQGVIGS 3293
QY 493 ---A-A-GKV---G---ASAS---D---GS-V-C--- 508
Db 3294 GCGOLAPYAHGDSLYFNGCQIRQA-ATKPLDLTRASKIMFVLOIGMSQTDSCNDSLSGP 3352
QY 509 ---VL---D---L-R-K 514
Db 3353 HAVDKAVLLQYVNVNGITWHVIAHQPKDFQAOQSVNVNPLEARMK 3399

RESULT 11

US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706

GENERAL INFORMATION:

APPLICANT: Leitner, Ernst
APPLICANT: Schneider, Elisabeth
APPLICANT: Schoergendorfer, Kurt
APPLICANT: Weber, Gerhard
TITLE OF INVENTION: Cyclosporin Synthetase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5827706artis Corporation
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,119A

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kaesenoiff, Melvyn

REGISTRATION NUMBER: 26,389

REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 503 8474

TELEFAX: 201 503 8807

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 15281 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Tolypocladium niveum

STRAIN: ATCC 34921

US-08-471-119A-2

Query Match 69.5%; Score 2550.8; DB 2; Length 15281;

Best Local Similarity 7.6%; pred. No. 8.2e-29;

Matches 468; Conservative 39; Mismatches 4; Indels 5662; Gaps 376;

QY 1 M-----S-----ISSDE-----VNF-----L-----V 12
Db 5261 MKATNELSSRYAAVLHI-SDEPLPIYKIDPEAWINFEGSRLTREALAQVLKENENAESV 5319
QY 13 -----Y-----R-----YL-----Q----- 17
Db 5320 AISNIPYKTVVERHIVRSLDQEDANAPESMDGSDWISAVRTRAQQCHTSLASDLFDIA 5379
QY 18 E-SGF-----S-----HS-----A-F-----T-----F----- 28
Db 5380 EDAGFVSVSWARQHSQHGALDAVFHHLKPAATESRVLIKFPYDTHQGRPLKSLTNQPLP 5439
QY 29 -----S-----S-----G-----I----- 31
Db 5440 AQSRRAELLIREGLQTLPPYIPISQITLIDRMLNANGKVDRELRARRAKITQSKPVE 5499
QY 32 -----S-----S-----H-----H-----IS----- 35
Db 5500 DIVPPRNSVEATVCKGFTDVLGVGEVITDNFFNLGHSIMATKLAARLQRLNTRISVRD 5559
QY 36 -----QSN-----I-----N-GA----- 42
Db 5560 VPDQPVVADLAAVIQNSAPHEPIKPADYTPVQPSQAGRLWFLDQLVGATWYLMPLG 5619
QY 43 -----L-----L-----V-----P-----P----- 46
Db 5620 IRLHSLRVDALATAISALEQRHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKD 5679
QY 47 -----A-----A-----S-----I----- 53
Db 5680 AYLAVLKHEQTTLFDLATEPGHVRVALIRLGEHEHLSIVMHHIISDCWSVEVLFDHMRH 5739
QY 54 -----Q-K-----GL-QY-----V-----E-----AEV----- 64
Db 5740 YSSALRQDDPMQILPLPIQYRDFAAWQKTEQVAEHQQLDYWTEHLADSTPAELLTDL 5799
QY 65 ---SI-----N-----ED-----G 70
Db 5800 PRPSILSGRANBELPLTIEGRHLKLRFCRVHQATPFFVILALRAAHVRLTGAEDATLG 5859
QY 71 T-----L-----F-----D----- 74
Db 5860 TPIANRPRELENMIGFFVNTQCMRIAEENDNFSLVRVRVSTATSAFANQDVPFESIV 5919
QY 75 -----G-----R-P-----I-----E-----S----- 80
Db 5920 SSLLPGSRDASRNPLVQVILAVHSQODLGKLFLEGRDEAVDSIASTRFDVEFHLFEHAD 5979
QY 81 -LS-----L-I-----D-----AVMP-----DV 91
Db 5980 RLSGSVLYAKELFKLRTIESVVSFLETURRALDQPLPLAVLPLTDTGCVGETASKGLLDV 6039

QY 92 -V-Q-TR-QQ-----97
Db 6040 PRTDPRDANIVEFQHVREATDAJAVKDATSILTVAQLDQSDRLAIWLSRRHMPET 6099
QY 98 -AY-----R-----D-KL-----103
Db 6100 LVGLAPRSCETIAMFGIMKANLAYPLDINSAPAARLSILSAVDGNKLVLLGSGVTAP 6159
QY 104 -A-Q-Q-----Q-----H-----107
Db 6160 EQENPEVAVGIOEILAGTCLDTQSNAPPSATSILAYVFTSGTGKPKGVVHERSVT 6219
QY 108 -A-----A-AA-A-A-AA-AAT-N-----119
Db 6220 RLAKPSNVISKLPGQARVAHLANIPDASIEIATLLNGATLVCLDYHTVLDLRTKEV 6279
QY 120 -Q-Q-----Q-----121
Db 6280 FERESITVVTLPALKQCVAEIPETLAHLDDLTYTGGDRVGHDAAMRARSCLKIGMSGY 6339
QY 122 G-----S-A-----K-----N-----GE-----128
Db 6340 GPTENTVITIEYDADEMFGVPIGKTVNSGAYVMORNQOLVPSGVVVELVVTGDGL 6399
QY 129 -N-TANGE-----134
Db 6400 ARGVTDPSLNKRFIYITVNGESIRAYRTGDRVYRPHDLQIEFFGRMDQOVKIRGHRIE 6459
QY 135 E-N-----GA-----138
Db 6460 PGEVESALLSHNSVQDAAVVICAPADQDSGAEMVAFVAARNTEDETOEBEAVDQVQGW 6519
QY 139 H-T-----I-ANN-----H-T-D-----147
Db 6520 THPETAAYSEVKDIROSEVGNDFMGWTSYDGSIEIDKTMHEWLNMTMILDAAREPGHV 6579
QY 148 -M-M-----E-V-----DG-----D 154
Db 6580 LEIGTGTGMFNLAKCPGLQGVGFPEPSKSAQFVNDAAQSPALKDGRSIVHVGATD 6639
QY 155 -VE-IPS-----N-----160
Db 6640 INKAGPIQRLVVINSVAQYFPTPEYLFRVVEALVQIPSVIRIVFGDMRTNAINRDFVAS 6699
QY 161 -KA-V-V-----LR-G-G-H-----168
Db 6700 RALHTLGEKANKRLVRQMIYELEANESEELTDPAPFTSURLGEKIKHVEILPKTMKAT 6759
QY 169 -E-S-E-V-F-----I-----174
Db 6760 NELSRYAAVLHVRSREQSTIHQVSPNAWIDFAADGLDRQTLINLLKEHKDAGTVAIG 6819
QY 175 -W-N-----P-V-SD-----A 176
Db 6820 NIPYSKTIVERFVNKSILSEDDMEGQNSLDGSAWAAVMAAQSCPSLDAMDVKETAOEA 6879
QY 177 -W-N-----P-V-SD-----182
Db 6880 GYQVEVSWARQWSONGALDAIFHHFPPKEGARTLIEFPTDYEGRNVTLTNRPLNSIOS 6939
QY 183 -LL-V-S-----G-----S-----188
Db 6940 RRLGTQIREKQLTLLPPYIPSRIMVLDQMPVNNNGKIDRKELVRAIVAPKPSAATRV 6999
QY 189 -G-----D-S-T-AR-----194
Db 7000 APRNEIETAILRDEFVIGTEVSVLDNFFDLGCHSLMATKLAARVSRRLDAHISIKDVPD 7059
QY 195 -I-----W-----N-----198
Db 7060 QPVLADLAASIORESAPHEIPORPYTGPAEQSAQGRWFLDQLNLGATWYLMPLAIRI 7119
QY 199 -S-----E-N-----ST-SG-----205

Db 7120 RGQLRVAALSAALFALERRHETLRTTFEESDGVGVQIVGEARNSDLRVHDSVGGDDGEVL 7179
QY 206 -P-----T-----207
Db 7180 EVLRREQVTFFDLSSSEPGHVRCLVKTGEEDHVLISVMHHIIYDGSVSDILRGELGQFYS 7239
QY 208 -Q-----L-V-LR-----H-----213
Db 7240 ALRGQDPLLANPLPIQYRDFAAWQREAKQVEEHQRLQVMSKQLVDSTPAELLTDLPRP 7299
QY 214 -C-----C-----I 215
Db 7300 SILSGRAGSVDTIEGSVYGALQSCFRSTRSVTTFFVLLTVFRIAHFRLTAVDDATIGTPI 7359
QY 216 -R-E-G-----G-----QDVP-----223
Db 7360 ANNRPELTLVGCFVNTQCMRISIAADDNFELVRQVRNVATAAYANQDVPFERIVSAL 7419
QY 224 -S-N-----K-----226
Db 7420 VPGSRNTRNPLVQLMFAVQSVEDYDQVREGLESVMMPGEASTRFDMEFHLVPGQKLT 7479
QY 227 -D-V-----TS 230
Db 7480 GSVLYSSDLFEQGTIQNFVDIFQECILRSVLDQPLTPIISVLFPFSNAISNLESLLDLEMP 7539
QY 231 -LD-----W-----233
Db 7540 DYPRDRTVVDLFREQAACPDIAVKDSSSULTYAQLDEQSDRVAALHERHMPAESLVG 7599
QY 234 -N-N-----S-EG-----237
Db 7600 VLSPRSCETIIAYFGIMKANLAYPLDLYVAPDARLAAILDVTEGERLLLLGAGVPOPGIQ 7659
QY 238 -T-L-----LA-----TG-----243
Db 7660 IPRLSYIAEALSHATTVDVTISIPQSAVSATSLAYVIFTSGTGKPGVMIEHRGIVRLVR 7719
QY 244 -S-----Y-----245
Db 7720 DTNVNVFPEGSGALPVSHFNSLAWDAATWIEYTAVLNGTTCIDRDTWLDIAALNSTFR 7779
QY 246 -D-----D-----246
Db 7780 KENVRAAFTPAFLKOCIAETPELVANLILHTAGDRDPGDANLAGKTAKGIFNVLGH 7839
QY 247 -G-F-----AR 250
Db 7840 TENTAVSTFYPVVEETFFVNGVGRGISNSHAYIIDRHQKLVPAQVGMGELILTGDOVAR 7899
QY 251 -I-W-T-K-DG-----256
Db 7900 GYTDSALNKDRFVYIDINGKSTMSYRTGDKARYPRDQGLEFFGRMDQMKIRGVRIEPG 7959
QY 257 -LG-----Q-----HK-----256
Db 7960 EVELTLLDHKSVAATVVVRPFPNGDPEMIAFTTDAEDDVQTHKAIYKHQILPAYMI 8019
QY 257 -N-LA-----ST-----261
Db 8020 PSHLVLDQMPVTDNGKVDKDLALRAQTQVKRSTAAARVPRDEVEAVLCEBYSNLEEV 8079
QY 262 -GPI-----FA-LKM-----N-KKG-----N-----279
Db 8080 EVGITDGFDLGGHSLLATKLAARLSRQLNTRVSVKDFDQPIADLADIIRRGSHRHP 8139
QY 267 -GPI-----FA-LKM-----N-KKG-----N-----279
Db 8140 IPATPYTGPEQSFAGRL-WFLEQLNGASVWLMFPFAIRMRGPLOTKALAVNALVHR 8198
QY 280 -F-----I-LS-AGVD-----K-TT-----290

Db 8199 HEALRTTFEDHGVGVQVIOFKSSQDLRIIDLSA-VDDTAYLAALKREQTAFDLTSEP 8257
QY 291 -----II-----
Db 8258 GWRVSLRLGLDDDDYILSVMHHSIDGWTVDVLRQELGQFYSAAIRGQEPQLSAKSLPIQ 8317
QY 293 -----W-----D-----294
Db 8318 YRDPVWQROENIQEQAQKLYWSQQLADSTPCEFLTDLPSPILSGEADAVPMVIDGT 8377
QY 295 -----AH-TG-----E-----299
Db 8378 VYQLLDFCRTHQVTSFVLLAAAFRTAHYRLTGTLDTAGTPIANRRNPBEGLEGIFVFN 8437
QY 300 -----A-QQ-----F-PF-----306
Db 8438 TOCRMALSETFESLVQVRLTTTEAFANQDVFFQIVSTLLPGSROTSRNPLVQVMF 8497
QY 307 -----H-----S-----308
Db 8498 ALQSQDLGRIQLEGMTDEALETPLSTRDLLEVHLFOEVGKLSGSLLYSTDLEFEVETIRG 8557
QY 309 -----A-PAL-----D-----V-314
Db 8558 IVDVLEILRRGLEQPKORLWAMPITDGTIKLRDQGLLTVAKPAYPRESSVIDLFRQOVA 8617
QY 315 -----D-W-----OSN-----N-T-----FA-SC-----325
Db 8618 AAPDAIAVWSSSTLTVADLDGOSNKLAWLQORNMAPETLVAVFAPRSCLTIVAFGLVL 8677
QY 326 -----C-----I-----HV-----C-334
Db 8678 KANLAYPLDVNAPARIEAILSAVPGHKLVIQAHGPELGLTMADELTVOIDEALASS 8737
QY 326 -----ST-----D-----M-329
Db 8738 SGDHEQIHASGPTATSLAYVMFTSGTGKPKGVIMDHRSIIRLVXNSDVVATLPTVRMA 8797
QY 330 -----C-----I-----HV-----C-334
Db 8798 NVSNLAFDISVQEIYTALLNGGTLVCLDYLTLDSKILYNVVFVEAQVNAAMFTFVLKQC 8857
QY 335 -----KL-----G-QD-----339
Db 8858 LGNMPALII SRLSVLFNVGDRDLDAHVAASGLQDAVYVAYGPTENGQWSTMYKVDVNEP 8917
QY 340 -----R-----P-----341
Db 8918 FVNGVPIGRSITNSGAYVMDGNQOLVSPGVNGEIVVTGDLARGYTDSALDEDRFVHVTI 8977
QY 342 -----IK-T-----P-O-----GH-----348
Db 8978 DGEENIKAYRTGDRVRYRPKDFIEFFGRMDQVKIRGHRIEPAEVEHALLGHDLVHDA 9037
QY 349 -----TN-----E-V-N-----A-IK-----356
Db 9038 VVLRKANQEPMEAFITISOEDETIEQESNKQVQGHGFHDVSRYADIKLOLSTFGHD 9097
QY 357 -----W-----D-P-----TG-NL-----363
Db 9098 FLGWTSMYDGVIPVNMKEWLOETTASLLDNRPFGHILEIGAGTGMLSNLKGVDGLQK 9157
QY 364 -----LA-----365
Db 9158 YVGLDPAPSAAI FVNEAVKSLPSLAGKARVLVGTALDIGSLDKNEIQPELVINSVAQYF 9217
QY 366 -----SC-----S-----D-----369
Db 9218 PTSEYLLKVVKAVVEPSVKRVFGDIRSQALNRDFLAARVALGDNASKEQIREKIAE 9277
QY 370 -----D-----M-T-----L-----K-374
Db 9278 LEBSEEBLLVDPAFFVLSRQLPNIKHIEVLKPLMKATNELSSRYAAVLHISHNEEQ 9337

QY 375 -----I-W-----S-----M-----K-----QD-381
Db 9338 LIQIDPTAWDFAATQKDSQGLRNLLQOGRDDVMIAVGNIPYSKTIIVERHIMNSLDQDH 9397
QY 382 -----381
Db 9398 VNSLDGTSWISDARSAAAICTSPDAPALTQLAKEGFRVELSWARQORSONGALDAVFHRL 9457
QY 382 -----NC-VH-D-----LQ-----388
Db 9458 ATDANCERSRVLVHFPTDHQGRQLRTLNRPLQARSRRIESQVFEALQTPALPAYMIPSR 9517
QY 389 -----388
Db 9518 IIVLPOMPTNANGKVDKQLARRAQVVAKRKAVSARVAPRNDTEIVLCEYADILGTEVG 9577
QY 389 -----388
Db 9578 ITDNPFDMGHSILMATKLAARLSRRLDTRVTKEVDFKPVLDLAASAIEQGSTPHLPAS 9637
QY 389 -----A-----H-----390
Db 9638 SVYSGPEQSYAQGRLWFLDQFNLNATWYHMSLAMRELLGFLNMDALDVALRALEORHETL 9697
QY 391 -----N-KE-----393
Db 9698 RTTFEAKDQIGVOVVEHAGMKRLKVLDSKNEKEHMAVLENEQMRPFTLASPEGMKGHL 9757
QY 394 -----I-YT-----I-----397
Db 9758 ARLGPTETIILSLVHMFMFSDGWSVDILRQELGQFYSAALRGDRDPLSQVKPLPIQYRDFAA 9817
QY 398 -----K-W-----S-P-T-GP-----404
Db 9818 WQKEAAQVAHERQLAYMENQADSTPGBLTDFPRPQFLSGKAGVIVPTIEGVPVYEKLL 9877
QY 405 -----GT-N-N-P-----N-----410
Db 9878 KFSKERQVTLFSVLLTAFRATHPRLTGAEDATIGTPIANRPELHIIIGFFVNTQCMRL 9937
QY 411 -----A-N-----LM-----414
Db 9938 LLDTGSTFESLVQHVRSVATDAYSNQDIPFERIVSALLPGSRDASRPLIQLMFALHSOP 9997
QY 415 -----L-A-----S-A-----S-F-420
Db 9998 DLGNITLEGLEHERLPTSVATRFDMEFHLFQEPNKLSGSILFADELFPQETINSVTVFQ 10057
QY 421 -----D-ST-----V-R-----425
Db 10058 EILRRGLDQPOVSIISTMPLTDGLIDLEKLGLEIESNPRDYSVVDVFRQOVAANPNAP 10117
QY 426 -----L-W-----427
Db 10118 AVVDSETSMYTSLDQKSEQIAAHLHAQGLRPESLIVMAPRSFETIVSLFGILKAGAY 10177
QY 428 -----DV-----DR-----431
Db 10178 LPLDVNSPAARIQPIILSEVEGKRLVLLGSGIDMPOSDRMDVETARIQDILTNTKVERSDP 10237
QY 432 -----431
Db 10238 MSRPSTSLAYVIFTSGSTRPKGVMEHNRNLRVVKQSNVTSQLPQDLRMAHISNLAFD 10297
QY 432 -----G-ICI-----H-----436
Db 10298 ASIWEIFTAILNGGALICIDYFTLLDSQALRTTFEKARVNATLFAFALLKECLNAPTFL 10357
QY 437 -----436
Db 10358 EDLKVLYIGGDRDLATDAAKIQALVKGTVYVAYGPTENTVMSTIYRLTGESYANGVPIC 10417

QY 437 -----TL----- 438
Db 10418 NAVSSGAYIMDKORLVPVGMGELVSGDLGRTNSTLANDRFVDIVINDOKARAY 10477
QY 439 -----T-----K--HQ-EP----- 444
Db 10478 RTGDRTRYRKPQSGIEFFGRMDQOVKIRGRHVEPAEVEQAMLGKNAIHDAAVVQVAVDQG 10537
QY 445 -----V-----Y-----S-- 447
Db 10538 ETEMIGFVSNASDRFSEGEBEITNOVQEWEDHFEFESTAYAGIEAIDQATLGRDFTSWTSMY 10597
QY 448 -----V-----A-----F-----S--P-- 452
Db 10598 NGNLIDKAEMEELDDTMSQLDKEDARPCEAIEGTGTGMVLFNLPKNDGLGESYVGIPEPSR 10657
QY 453 -----D-----G-----R-YL--A 458
Db 10658 SAALFVDKAAQDFGLGKQTQIILVGTAEIDIKLVKDFHPDVVVINSVAQFPSPRSYLVOIA 10717
QY 459 S-----G--S-----F-----DK----- 464
Db 10718 SELHMTSVKTIFFGDMRSWATNRDFLVSRALYTLGDKATKQIQREVARLENEDELLV 10777
QY 465 -----W-----NT-----C-----VH-I 468
Db 10778 DPAFTSLTSQMPGVKVEILPKRMRTSNELSSRYAAVLHICRDGEGNRYGRRVHSV 10837
QY 469 -----W-----T-----S-----Y----- 471
Db 10838 EENAWIDFASGMDRHALVQMLDERRDAKTVAIGNIPHSNTINERHFTTSLDTEGEGIAQ 10897
QY 472 -----Q-----T-----S-----GAL--V--H-- 478
Db 10898 DSLDGSAMQSAKAMARCPCLSVTELVEIGQAAGFRVEVSWARQSRQSGALDVVFHLE 10957
QY 479 -----S-----Y----- 480
Db 10958 DDRVGRVLNFPDFERLPBPGTGLSRPLQRIQNRPFESQIREQLQTLPLPYVWPSRIVV 11017
QY 481 -----R-----G-T----- 483
Db 11018 LERMPNANSKVDKELARKARTLOTIKPSATRVAPRNDIEAVLCDFQAVLGVTVGVMD 11077
QY 484 -----G-----R-----E-----B----- 488
Db 11078 NPFELGSHLMATKLAARLSRRLDTRVSVKDIENQPIQLADVQVOTGAPHEAIPSTPY 11137
QY 489 -----VC-----W-----NA-----A-----G----- 495
Db 11138 SGPEQSFSGRLWFLDQLNLNASWYHMPPLASRLRGLRTEALQSLATIEARHESLRTT 11197
QY 496 -----D-----K-----V-G-----A--SA----- 502
Db 11198 FEBQDGVVQIVRAARKQLRIIDVSGTEDAYLAALKQEODAFDLTAEPGMRVALLRLG 11257
QY 503 -----SDG-SV----- 507
Db 11258 PDDHVLIVWHHIIISDGSVDILRQELGQLYSNASSQAPLPIQYRDPFAIWQKDSQIAE 11317
QY 508 -----L-----D-----LR--K 514
Db 11318 HQQLNWKQLVNSKPAELLADFTPRKALSGDADVIPIBIDQVYNLRSFCRARHVT 11377
QY 509 -----L-----D-----LR--K 514
Db 11378 FVALLAAFRAAHYRLTGAEDATIGSPIANRNRPEGLGICFVNTQCLRIPEK 11430

RESULT 12

US-09-052-469-8
; Sequence 8, Application US/09052469
; Patent No. 6380360

GENERAL INFORMATION:
APPLICANT: Harris et al.
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: One Financial Center
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk, 3.50 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/052,469
FILING DATE: Concurrently herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/422,582
FILING DATE: 14-April-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9507766.5
FILING DATE: 13-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9411900.5
FILING DATE: 14-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326470.3
FILING DATE: 24-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Ph.D., Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/74165
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 345-9100
TELEFAX: (617) 345-9111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-052-469-8

Query Match 69.5%; Score 2548.4; DB 3; Length 4302;
Best Local Similarity 9.8%; Pred. No. 1e-11;
Matches 412; Conservative 83; Mismatches 13; Indels 3676; Gaps 357;

QY 1 MSIS-----S-----D-----EVN-----F----- 10
Db 72 LDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEGIFANLNLSEINLGNPPECDC 131
QY 11 -LVY--RY-----L--Q-----E-----SG-- 20
Db 132 GLAWLPRWAEQQVRYVQPEAAATCAGPGSLAGQPLLGIPLLDSCGGEYVACLPDNSSGT 191
QY 21 ----PS--H-----SAF-----I----- 26
Db 192 VAAVFSAAHEGLLOPEACSAFCFTGQGLAALSEQWCLGAAQPSASFACLSLCSGP 251
QY 27 -----T-----F-----G-----I----- 30
Db 252 PPPAPTCTGPTLLQHVFPASPATGATLVGPHGPLASQLAAPHIAAPLPTVTRWDFDGS 311
QY 31 -E-----SH-----I-----S-QSN-- 38
Db 312 AEVDAAGPAASHRYVLPGRHYVTAVLALGAGSALLGTDVQVEAAPAAALVLCPPSSQSD 371

QY 39 -I-N--G--AL--V--P--P--P-- 46
Db 372 SLDSIQNRGSGLEAAYSIVALGEEPARVHPLCPDSEITFONGHCYLVVEKAWLQ 431
QY 47 -AA--LIS--I--IQ-K--G--L-- 57
Db 432 AQEQCAWAGALAMVDSPAVQFLVSRVTRSLDMWIGFSTVQGVGEPAPQGEAFSLES 491
QY 58 Q--YV-E--AE--V--S-- 65
Db 492 CONWLPGEHPATAEHCVRLGPTGMCNTDLCSAPHYSVCELPQGGPVQDAENLLVGA PSG 551
QY 66 -IN--EDG--TL--F-- 73
Db 552 DLOGPLTFLAQDGLSAPHEPVEVMVPPGLRLSREAPLTTAEFGTOELRRPAQLRLQVYR 611
QY 74 -D--GR--P--I-- 78
Db 612 LLSTAGTPENGSEPSRSPDNRTQLAPACMPGGRWCEGANICLPDASCHPQACANGCTS 671
QY 79 -E--S--LSLI--DAV-M-P-DVV-- 92
Db 672 GPGLPGAPYALWREFLFSVAAGPPAQS-VTLHGQD-VLMLPGDLVLQHDAGPGALLHC 729
QY 93 -Q--T--R--Q-- 96
Db 730 SPAPGHPGPOAPYLSANASSWLPHLPALQEGTWACAPACALRLLAATEQLTVLLGLRPNPG 789
QY 97 -QA--Y--RD-KL--A--Q-- 105
Db 790 LRMPGRVEBRAEVGVRHNLSCSPDVVSPVAGLVIYPAPRDGRLYVPTNGSALVLQV 849
QY 106 -OH-- 107
Db 850 DSGANATATARPFGSVARSFENVCPALVATFVPGCPWETNDTLFSVALPMLSEGEHV 909
QY 108 -A--A--A--A--A--A-- 113
Db 910 DVVENSASRANLSLRTAEPTICGLRATPSPEARVLQGVLYRSPVVEAGSMVFRWTI 969
QY 114 -AA--A--AT--N-Q-Q-- 121
Db 970 NDQSLTFQNVFNVIQSAAPKLSLTASNHNVSNTVNTVVERNMNRQGLQVSTVPA 1029
QY 122 -C--SA-K--N--GE--N--TA-- 131
Db 1030 VLSFNATLVLTGVLVDSAVEAFLWNFGDGEQALHQFQPPYNESFPVDPSPVAQLVEH 1089
QY 132 N--GE--EN--G--A--H 139
Db 1090 NVMTYAAPGEYLLTVLASNAFENLTQOVVSVRASLPSVAVGSDGVLVAGRPTVFPYH 1149
QY 140 -T--I--ANNHT-- 146
Db 1150 PLSPGGLVLTWDFGDSPLVLTQSPAANNHTYASRGTYHVLRVLEVNNTVSGAAAQADRVF 1209
QY 147 -DM-M-E--V--D-GD-- 154
Db 1210 EELRGLSVDMSLAVEQCAPVVVSAAVQTGDNITWTDFMDGDTGLSGPEATVEHVYLRQ 1269
QY 155 -VS--IPS--NKA-- 162
Db 1270 CTVTVGAASPAHGLARSLHLVFLVLRVEPAACTPQDARLTAYVTCNPAHYLFDWT 1329
QY 163 -V--V--L-R-G--H--ESEV-- 172
Db 1330 FGDSSNTTVRGCTVTHNTRSGTPELALVLSRVNRAHYFTSICEPEVGNVTLQPER 1389
QY 173 -FI--CAW--NP-- 179
Db 1390 QFVQGDGEAWLVACAWPPFPYRYTWDFTGTEAAPTARGPEVTFIYRDPGSLVTVTASN 1449

QY 180 -VS--D--L--LV-S--GS-- 190
Db 1450 NISAANDSALVEQEPVLTSTIKVNGSLGLELOQPYLFSAGVGRPASVYLDLGDGWL 1509
QY 191 -S--TARI--WN-L--SE--NST--SG-- 206
Db 1510 GPBVTTHAYNSTGDTFVRVAGWNEVSRSSEAWLNVTKRRVRGLVVNASRTVPLNGSVSFS 1569
QY 207 T--QL--VL--RHC--I--RE-G--G--QD-- 221
Db 1570 TSLEAGSDRYSVNLCDR-CTPIPGGTISYTFRSVGTNIIIVTAENEVSQAQDSIFVYV 1628
QY 222 -V--PSN--KD--V--T-- 231
Db 1629 LQIEGLQVVGGRYPFTHVTOLOQAVVRDCTNVSYSWTAWDRGPALAGSGKFSLTVL 1688
QY 232 -D--W--N-- 235
Db 1689 EAGTYHVQLRATNMLGSAWADCTMDFVEPVGLMWLTASPNPAAVNTSVTLASLAGSGV 1748
QY 236 -EG--T--L-LAT--GS--YD-- 246
Db 1749 VYTWSLEEGSLWETSEPTTHSPTPLGLHLVTMTAGNPLGSANATVEVDQVPVSGLSIR 1808
QY 247 -G-F-A--R--I--W--TK--D-G--NL 258
Db 1809 ASEPGGSFVAAGSSVPFGQLATGNTVSWCNAVPGSSKRGPHVTWVPDAGTFSIRLN 1867
QY 259 AS--T--L--QO--H--K--G-- 267
Db 1868 ASNAVSWSATYNTLTAEPIVGLVLWASSKVAPQOLVHFQILLAAAGSAVTFRLQVGGAN 1927
QY 268 P-I--F--AL--K--W--N--KKG-- 278
Db 1928 PEVLPGREFSHSPRVGDHVSVRGKXHVSAQAQVRIIVLEAVSGLQMPNCEPGIATG 1987
QY 279 -NF--ILS--AG-- 285
Db 1988 TERNFTARVORGRVAYANYFSLOKVGDSLVLSGRDVTYTPVAGLLEIQVRAFNALG 2047
QY 286 -V-D--K--T--T--II--WD-- 294
Db 2048 SENRTLVLVEQDAVQYVALQSGPCFTNRSAQFEAATSPSRPRVAYHWDFGDGPQDQDE 2107
QY 295 -A-HT--G--E--A--KQ--Q--FP-- 305
Db 2108 PRAEHSYLRPGDYRVQVNASNLVSFFVAQATVTQVLACREBEVDVVLPLQVLMRRSQ 2167
QY 306 F--H--SA--P--AL--DV-- 314
Db 2168 YLEAHVDLRDCVTYQTEYRWEVYRTASCORPGRPARVALPGVDVSRPLVLPRLALPVGH 2227
QY 315 -D--D--W--QS--N-- 319
Db 2228 YCFVFWVSGDPLTQSIQANVTVAPERLVPPIEGGSYRVWSDTRDLVDGSESYDNLE 2287
QY 320 -N--TF-- 322
Db 2288 DGDQTPLSFWHACVASTQREAGCALNFGPRGSSTVTIPRERLAAGVEYTFSLTVWKAGR 2347
QY 323 -A--SC-- 325
Db 2348 KBEATNTVLIRSGRVPVLSLVCVSCQAQVYEVSRSSVYVLEGRCLNCSSGSKRGWAA 2407
QY 326 -S--T--D--M--C--I-- 331
Db 2408 RTFSNKTLLVLDETTSTGSGAGMRLVLRGVRDGEYTFTLVLRSGBEEGCASIRLSP 2467
QY 332 -H--V--C-- 334
Db 2468 NRPPGSGCRLPPLGAVHALTTKVHFECTGWHDAEDAGAPLVYALLRLRCRQGHCEFCV 2527
QY 335 -K--L--G--QD--R--P-- 342

Db 2528 YKSLSSYGAVLPGRPHFVGLVAVVQDQGAVALNRSIAITLPEPNGSATGLTVM 2587
QY 343 ---KT---FG---H---T-NE---V---N- 353
Db 2588 LHGLTASVLPGLLRQDPQHVIBYSALVTLNVEYERDALDVAABPKHERQRAIRKNIT 2647
QY 354 ---LA---S---C---SD--- 356
Db 2648 ETIVSLRVHTVDIIOIAAALACQMGPSRELVCRLKQTLHKLEAMMLILOAETAGTV 2707
QY 357 ---WD---PT-G---NL--- 363
Db 2708 TPTAIGDSILNITGDLHLASSDVRAPQPSSELCAESPSRMVASQAYNLTSALMRILMRSR 2767
QY 364 ---LA---S---C---SD--- 369
Db 2768 VLNEEPLTLAGEEIVAQGRSDPRSLCYGAPGCHFSIPEAFSGALANLSDVVQLIF 2827
QY 370 -D- -M- -T- -L- -KI- -W- - 376
Db 2828 LVDSNFPFGYISNYTVSTKVASKAFQTQGAQIPIERLASERAITVKVPNNSDWAARGH 2887
QY 377 ---SM---KQ---D--- 381
Db 2888 RSSANSANSVVQPQASVGAVVTLDSNPAAGLHLQNLNYTLDDGHVLSSEPEPYLAVLH 2947
QY 382 ---NC---V---H---DLOA---H-N---K---EI 394
Db 2948 SEPRNEHCASRRIRPESLOGADHRPYTFFISPGSRD-PAGSYHLNLSHFRWSALQV 3006
QY 395 ---YT---IKW---SP---T-G--- 403
Db 3007 SVGLYTSLOQYFSEEDMV-WRTGGLPLLETSRQAVCLTRHLTAFGASLFPVPPSHRVFV 3065
QY 404 -P- -GT- -N- 404
Db 3066 FPEPTADVNIYMLTCAVCLVYVMAAILHLKLDQDASGRRAIPFCGQGRPKYELVK 3125
QY 405 ---GT---N- 407
Db 3126 TGMGRSGTTHAVGIMLYGVDSRSGRHLDDGAFHNSLIDIFRIATPHSLGSKIRVM 3185
QY 408 -N- -P- -N- -AN- -L- -M- -L- -AS- - 419
Db 3186 HDNKGSLPAWFLQHVIVRDLQARSAPFLVNDWLSVETEANGVLKEVLAASDAALLRP 3245
QY 420 ---PD---S-T-V-R---L---W---D- 428
Db 3246 RRLVLAELQGFQKHILWSIWDPRPSRTRIQATCCVLLICLFLGANAVWYGAAGDS 3305
QY 429 ---VD---R---G--- 432
Db 3306 AYSTGHVSRSLPSLSDTVAVGLSVVVPVYLAILFLFRMSRSKVAGSPSPAGQVQL 3365
QY 433 -I- -C- -I- -H- -TLT- 439
Db 3366 DIDSCLDSSVLDSSFTFSGLHAEAFVGMKSDLFLDDSKLVCWPSGEGTSLWPDLSD 3425
QY 440 ---K- 440
Db 3426 PSIVGSNLRLQARGHGLGPEEDGFSLASPYSPAKSPASDEDLIQVLAEGVSSPAP 3485
QY 441 ---H- -Q- -E- - 443
Db 3486 TQDTHMETDLSLSTPGKTTETLALQRLGELGPPSPGLNWQQAARLSRTGLVEGLR 3545
QY 444 ---P- -V-Y- -SVA- 449
Db 3546 KRLLPAWCASLAHGLSLLLVAVAVSGWVGASFPFGVSVALLSSASFLASFLGWEP 3605
QY 450 ---PS- -PD- 453

Db 3606 KYLLEALYFSLVAKRLHPDEDDTLVESPAVTPVSARVPRVRPHGFALFLAKEARKVKR 3665
QY 454 -G- -R- -Y- -LAS- -G- -S- -F- - 462
Db 3666 LHGMLSLLVYMLFLLVTLASYGDSCHGHAYRLOSATKQELHSAFLAITRSEELW 3725
QY 463 ---DK- -C- -V- - 466
Db 3726 MAHVLLPYVHGNQSSPELGPRLRQVRLQEAALYDPDPPGPRVHTCSAAGGFSSTSDYDV 3785
QY 467 -H- -I- -W- -N- - 470
Db 3786 SPHNGSGTWASAPDILGAWSGSCAVYDGGVQBELGSLSEBSRDLRLFLQHLNMLDNR 3845
QY 471 ---TQ- -T- -G- -AL- - 476
Db 3846 SRAVLELTRYSPAVGLHAAVTLRLEPPAAGRALAALSVRPPALRLSAGLSPLLTVC 3905
QY 477 ---VH- -SY- -R- -G- -T- - 483
Db 3906 LLLFAVHFAVAEARTWHREGRVRLGAWARWLLVALTAATLVALQGLGAADROWTRF 3965
QY 484 -G- -G- -IP- -E- - 488
Db 3966 VRGRPRRFTSFDQVAHVSSAARGLAALLFLLLVKAQAQHVRFVROMSVFEGKTLRALPEL 4025
QY 489 -V- - 489
Db 4026 LGVTGLVGLVAYAOALAILVSSCVDSLWSVAQALLVLCPTGLTCLPAESWHLSPLL 4085
QY 490 -C- -WNA- -A- -G- -D- - 496
Db 4086 CVGLMALRLWALRGAVILRWYHALRGELYPAMEPDYENVELFLRLRLMGLSKV 4145
QY 497 -KV- -G- -A-S- -AS- -DG-SV- - 507
Db 4146 KEPRHKVRPEGMPLPSRSRSGSKVSPDVPFPSPAGSDASHPSSTSSQLDGLSVSLGR 4205
QY 508 -C- -V- -L- -D- -L- -RK- 514
Db 4206 RCEPEPSRLOAVFEALLTOFDRLNQATDVEDYOLEQOLHSLQGR 4249

RESULT 13

US-08-422-582-8
; Sequence 8, Application US/08422582
; Patent No. 6485960
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 75 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,582
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9507766.5
FILING DATE: 13-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326470.3
FILING DATE: 24-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Ph.D., Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/53313 (MRC-006xx)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 345-9100
TELEFAX: (617) 345-9111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-422-582-8

Query Match 69.5%; Score 2548.4; DB 4; Length 4302;
Best Local Similarity 9.8%; Pred. No. 1e-31;
Matches 412; Conservative 83; Mismatches 13; Indels 3676; Gaps 357;

QY 1 MSIS-----S-----D-----EYN-----F-----10
DB 72 LDVSHNLLRALDVGLLANLSALAEIDISNNKISTLBEGIPANLFNLSGNPPECDC 131
QY 11 -LVY--RY-----L--Q-----E-----SG--20
DB 132 GLAWLPRWAEQQRVVQPPAATCAGPGSLAGPLGILLDSCGEEYVACLDPNSSGT 191
QY 21 ----FS--H-----SAF-----26
DB 192 VAAVSFAAHEGLQPEACSAFCSTGQGLAALSEQGWLCLGAAQSSASFACLSLCSGP 251
QY 27 ----T-----F-----G-----I-----30
DB 252 PPPAPTCTGPTLLQHVFPASPAGATLVGPHGLASQOLAFAHIAFLPVTATRWDFGDS 311
QY 31 -E-----SH-----AL-----V-----P-----I-----S-QSN--38
DB 312 AEVDAAGPAASHRYVLPGRVHTAVLALGAGSALLGTDVQVEAAPAALELVCFSSVSDE 371
QY 39 ----I--N--G-----AL-----V-----P-----I-----Q--K-----G-----L--57
DB 372 SLDLISQNRGGSGLEAAYSIVALGEBPARAVHPLCPSTDEIFPGNGHCYRLVVEKAAMLQ 431
QY 47 ----AA-----LIS-----I-----IQ--K-----G-----L--57
DB 432 AQEQCAWAGALAMVDSPAVQRFVRSVTRSLDWMIGFSTVGQVEGPAQGEAFSLBS 491
QY 58 -Q-----IN-----EDG-----YV-E-----AE--V--S--65
DB 492 CQNLWLPCEHPATAEHCVLGTGNCNTDLCSPHVSVCLOPGGPVQDAENLLVGAPSG 551
QY 66 ----IN-----EDG-----YV-E-----AE--V--S--65
DB 552 DLQGLTPLAQDGLSAPHEPVEVMPFGLRLSREAFLLTAETGQELRRPAQLRLQVVR 611
QY 74 ----D-----D-----CR--P--I-----78
DB 612 LLSTAGTPENGSEPSRSPDNRQTOLAPACMPGGRCWCPGANICLPDASCHPQACANGCTS 671
QY 79 -----E--S-----LSLI-----DAV-M-P-DVV-----92
DB 672 GPGLPGAPYALWREFLFSVAAGPPAQS-VTLHGQD-VLMLPGDLLVGLQHDAGPGALLHC 729
QY 93 -----Q-----T-----R-----Q-----96
DB 730 SPAPGHGPPQAPYLSANASSWLPPLPAQLGEGTWACPAALRLMLAATEQLTVLLGLRPNPG 789

QY 97 ----QA-----Y-----RD-KL-----A--Q--105
DB 790 LRMPGRYEVRAEVGNGVSRHNLSCSFDVVSVPAGLRIYIPAPRDGRLYVPTNGSALVLQV 849
QY 106 -----A-----A-----A-----A-----OH--107
DB 850 DSGANATATARWPGGSVSARFENVCPALVATFVPGCPMETNDTLFSPWALPMLSEGEHV 909
QY 108 -----A-----A-----A-----A-----A-----113
DB 910 DVVENSASRANLSLRVTAEPIICGLRATPSPEARVLQGLVRYSPVWEAGSDMFRWTI 969
QY 114 -----A-----A-----A-----A-----AT--N--Q--Q-----121
DB 970 NDKQSLTFQNVVFNVIYQSAANVFKLSLTASNHSNVTVNVTVERMRNMQGLQVSTVPA 1029
QY 122 -----G-----SA-K--N--GE-----N-----TA-----131
DB 1030 VLSPNATLVLTGGVLVDSAVEVAFLWNFGDGEALHOFPPYNESFPVDPSPVAQVLVEH 1089
QY 132 N-----GE-----EN-----G-----A-----H--139
DB 1090 NVMHYAAPGEYLLTLVLASNAPENLTQQVPVSVRASLPVAVGSDGLVAGRPVTFYPH 1149
QY 140 -----T-----I-----ANNHT-----146
DB 1150 PIPSPGGVLYTWDGPGSPVLTSQSPAAHNTYASRGTYHVRLEVNNTVSGAAQAQVRVF 1209
QY 147 -----DM-M--E-----V-----D-GD-----154
DB 1210 EELRGLSVDMSLAVEGAPVWVSAAYQTGDNTITWTFDMDGDTVLSGFEATVEHVYLAQN 1269
QY 155 -----VE-----IPS-----NKA-----162
DB 1270 CTVTGGAASPAGHARSLHVLVLEVRVEPAACIPTQDARLTAYVTGNPAHYLPDWT 1329
QY 163 -----V-----V--L-R-G-----H-----ESEV-----172
DB 1330 FGDGSSNTVRCGPTVTHNFTSGTFPLALVLSRVNRAHYFTSICVEVEGVNVTLOPER 1389
QY 173 -FI-----CAW-----NP-----179
DB 1390 QFVQLGDEAWLVACAWPPPPRYTWDFTGTEAAPTRARGPEVTFIYRDPGSLVTVTASN 1449
QY 180 -VS--D--L-----LV-S--GS-----GD-----190
DB 1450 NISAANDSALVEQEPVLVTSIKVNGSLGLELQOQPYLFSAGVGRPASVYMLDGLDGMLE 1509
QY 191 -----S-----TARI--WN-L--SE--NST-----SG-----P-----206
DB 1510 GPEVTHAYNSTGDFTVRVAGWNEVSRSEAWLVTVKVRGLVNVNASRTVVPPLNGSVSFS 1569
QY 207 T-----OL--VL--RHC-----I-----RE-G-----G--QD-----221
DB 1570 TSLEAGSDVRYSWVLCDR-CTPIPGGPTISYTFRSVGTENIIIVTAENVEGSAQDSIFVYV 1628
QY 222 -----V-----PSN-----KD--V--T-----SL-----231
DB 1629 LQLIEGLVVGGRYFPTNHTVQLQAVRDGNTVSVSWTAWDRDGPALAGSGKGFSLTVL 1688
QY 232 -----D-----W-----N-----S-----235
DB 1689 EAGTHVQVLRATNMLGSANADCTMDPVEVPVGLMVTASPNPAAVNTSVLSAELAGSGV 1748
QY 236 -----EG-----T--L--LAT-----GS-----YD-----246
DB 1749 VYTWLSEGLSWETSBPFTTHSGFTPLGLHLVTWAGNPLGSANATVEVDVQVPVSGLSIR 1808
QY 247 -----G-F-A-----R-----I--W-----TK--D-G-----NL--258
DB 1809 ASEPGGSFVAAGSVFVFWGQLATGNTVSWCWA VPGSSKRGPHVTWTFPDAGTFSIRLN- 1867
QY 259 AS-----T-----L-----GQ--H-----K-----G--267

Db	1868	ASNAVSVSATYNLTAEPIVGLVWASSKVAPQGVHVFQILLAAAGSAVTFRLQVGGAN	1927	Db	2948	SBPRPNEHNCASRRIRPESLOQADHRPYTFTISPSGRD-PAGSYHLNLSHFWRWSALQV	3006
Qy	268	P-I- -F--AL- -K--W- -N--KG- -	278	Qy	395	---YT- -IKW- -SP- -T--G- -	403
Db	1928	PEVLPGRFSPRGDHHVSVRGKHHVSAQAQVRIVVLEAVGLOMPNCCPEGIATG	1987	Db	3007	SVGLYTLSCQYFSEEDMV-WRTEGLLPLEETSPRQAVCLTRHLTATFASGLFVPPSHVRFV	3065
Qy	279	---NF- -ILS- -AG- -	285	Qy	404	---P- -	404
Db	1988	TERNFTARVQGRSVAYAVFYSLQKVGDSLVLSGRDVTYTFVAAGLLEIQVAFNALG	2047	Db	3066	FPEPTADVNIYVMLTCAVCLVTYVMAAILHLKLDQDASGRRAIFPCGGRGPKYBILVK	3125
Qy	286	---V-D- -K--T- -T- -II--WD- -	294	Qy	405	---GT- -N- -	407
Db	2048	SENRTLVEQDAVQVVALQSGPCFTNRSQAFAATSPSPRVAYHWDGDSGPGQDTDE	2107	Db	3126	TGWRGSGTTAHHVIMLYGVDSRSGHRHLDGDPHFRNSLDIFRIATHPSLGSVMKIRVM	3185
Qy	295	---A-H- -G--E--A- -KO--O- -FP- -	305	Qy	408	---N--P- -N- -AN--L--M--L-AS--	419
Db	2108	PRAEHSYLRPGDVRVQVNASLVSFVFAQATVTTVQVLACKREPEVDVVLQVLMRRSQRN	2167	Db	3186	HDKGLSPAFLQHVIVRDLQTARSFAFFLVNDWLSVETANGGLVEKEVLAASDAALLRF	3245
Qy	306	F--H- -SA- -P--AL--DV- -	314	Qy	420	---FD- -S--T--V--R- -L--W--D- -	428
Db	2168	YLEAHVDLRDCVYQTEYRWEVYTASCQPGPARVALPGVDVSRPRLVLPALPVGH	2227	Db	3246	RRLVVAELQRFQKHIMLSIWDPRPSRFTRIQATCCVLLICLFLGANAVMYGAVGDS	3305
Qy	315	---D- -N- -W- -QS--N--	319	Qy	429	---VD- -R- -G--	432
Db	2228	YCFVSVSGDTPLTQSIQANVTVAPELVP1IEGGSYRWSDRLVLDGSESDPNLE	2287	Db	3306	AYSTGHVSRSLPSVDTVAVGLVSSVYVYVYLA1LFLFRMSRSKVAGSPSPAGQVQL	3365
Qy	320	---N- -TF- -	322	Qy	433	-I--C- -I- -H- -TLT- -	439
Db	2288	DGQDTPLSFHWACVASTQREAGCALNFGPRGSSVTI1PRERLAAGVEYTFSLTVWKAGR	2347	Db	3366	DIDSLDSSVLSSFLTFPSGLHABAFVGMQKSDLFDDSKSLVCWPSGSGT1SWPDLSD	3425
Qy	323	---A- -SC- -	325	Qy	440	---K- -	440
Db	2348	KEBATQTVLIRSGRPIVLSVCSCKAQAVYEVSRSSVYLEGRCLNCSSGSKRGWAA	2407	Db	3426	PSIVGSNLQRLARGQGHGLGPEEDGFSLASPYSPAKSPFASDEDL1QQVLAEGVSSPAP	3485
Qy	326	---S--T--D- -M- -C--I--	331	Qy	441	---H- -O- -E--	443
Db	2408	RTFSNKTVLDETTTSGSAGMLVLRGVRGDEGVTFTLVLRSGSEEGCASIRLSP	2467	Db	3486	TQDTHMETDLSLSSSTPGEKTETLALQRLGBLPPSPGLNMEQPOAARLSRTGLVEGLR	3545
Qy	332	---H- -V- -C--	334	Qy	444	---P- -V--Y- -SVA- -	449
Db	2468	NRPLGSGCRLPPLGAVHALTTKVHFECTGWHDAEDAGAPLVYALLRLRCRQGHCEFCV	2527	Db	3546	KRLLPAMCASLAHGLSLLLVAVAVAGVWGASFPFGVSVAMLLSSASFLASFLGWEP	3605
Qy	335	---L- -G--OD- -R- -P- -I--	342	Qy	450	---PS- -	453
Db	2528	YKGLSSYGAIVLPPGRPHFVGLVAVVQDLGAAVVALNRSIAITLPEPNSATGLTVW	2587	Db	3606	KVLLLEALYFSLVAKRLHPDEDDTLVESPAVTPVSARVPRVRPHGFALFLAKEARKVR	3665
Qy	343	---KT- -FOG- -H- -T--NE- -V- -N--	353	Qy	454	---G--R--Y- -LAS- -G- -S--F- -	462
Db	2588	LHGLTASVLPGLLRQDPQHVIYSALVTVLNEYERALDVAAEPKHERQHQRAIRKNIT	2647	Db	3666	LHGMLRSLVYMLFLVLTLLASYGDSACHGAYRLQSAIKQELHSAFLAITRSEELWPW	3725
Qy	354	---A- -I--K- -	356	Qy	463	---DK- -C- -V--	466
Db	2648	ETLVSLRVHTVDDIQIAAALAAQCMGFSRELVCRLKOTLHKLEAMML1QOATTAGTV	2707	Db	3726	MAHVLLPYVHGNQSSPELGPPLRQVRLQOALYPOPPGPRVHTCSAAGFSTSDYDVWE	3785
Qy	357	---WD- -PT--G- -NL- -	363	Qy	467	---H- -I- -W--N--	470
Db	2708	TPTAIGDSILNITGDLIHLASSDVRAPOFSELGAESPMSVQANLTSALMRLMRSR	2767	Db	3786	SPHNGSGTWASAPDLLGAWNSGCAVYDSGGVQELGSLSESRDLRFLQLHNLNDR	3845
Qy	364	---LA- -S- -C- -SD- -	369	Qy	471	---TQ- -T- -G- -AL- -	476
Db	2768	VLNEEPLTLAGEEIVAGKESDRSLLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIF	2827	Db	3846	SRAVLELTRYSPAVGLHAAVTLRLEFPAGRALAALSVRFPALRRLSAGLSLPLTTSVC	3905
Qy	370	---D- -M--T- -L- -KI- -W- -	376	Qy	477	---VH- -SY- -R- -G--T--	483
Db	2828	LVDNSPPFPGYISNYTVSTKVSMAPQOAGAPIERLASERAITVKVPNNSDWAARGH	2887	Db	3906	LLLFAVHFVABARTWHREGWRVRLGAWARWLLVALTAATVALVLAOLGAADROWTRF	3965
Qy	377	---SM- -KO- -D- -	381	Qy	484	---G- -G- -IF- -E--	488
Db	2888	RSSANSANSVVQFQASVAGAVTLDSSNPAAGLHLQNLTYLLDGHVLSPEPEYLAIVLH	2947	Db	3966	VGRPRRFTSPDOVAHVSSAARGLAASLLFLLLVKAAQHVRFVRQWSVFGKTLCRALPEL	4025
Qy	382	---NC- -V- -H- -DIQA- -H--N--K--E--	394	Qy	489	---V- -	489
				Db	4026	LGVTGLGVVAVAGLA1LLVSSCVDSLWSVAQALLVLCPGTGLTCLCPAESWHLSPLL	4085

QY 490 C-----WNA-----A-G-----D----- 496
Db 4086 CUGLWALRLMGLALRGALVILRWYHALRGELYRPAWEPQDYEMVELFLRLRLMMGLSKV 4145
QY 497 -----KV-----G-----A-S-AS-----DG-SV----- 507
Db 4146 KEFRHKVRFEGMEPLPSRSGKSVSPDVPVPPSAGSDASHPTSSSQDGLSVSLGRIGT 4205
QY 508 -C-----V-----L-----D-----L--RK 514
Db 4206 RCEPEPSRLQAVFEALLTQFDRLNQATEDVYQLEQLHLSLQGR 4249

RESULT 14

US-09-052-262-8
; Sequence 8, Application US/09052262
; Patent No. 6656681
; GENERAL INFORMATION:
; APPLICANT: Harris et al.
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,262
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74118
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-052-262-8

Query Match 69.5%; Score 2548.4; DB 4; Length 4302;
Best Local Similarity 9.8%; Pred.No.1e-31;
Matches 412; Conservative 83; Mismatches 13; Indels 3676; Gaps 357;

QY 1 MSIS-----S-----D-----EVN-----F----- 10

Db 72 LDVSHNLLRALDVGLLANLSALAEIDIENNKTSTLEEGIFANLFLNLSEINLSGNPFECDC 131
QY 11 -LVY--RY-----L-Q-----E-----SQ- 20
Db 132 GLAWLPRWAEEOQVRVQPEAATCAGPSLAGOPLLGILLDSDGCGEYVACLPNSSGT 191
QY 21 ----FS-H-----SAF----- 26
Db 192 VAAVSFAAHEGLLOPEACSAFCFTGGQLAALSQGWCLCGAAQPPSSASFACLSICSP 251
QY 27 ----T-----F-----G-----I----- 30
Db 252 PPPAPTCKRGPTLLQHVFPASPGATLVGPHGPLASQLAFAHIAAPLPVTATRWDFGDS 311
QY 31 -E-----SH-----I-----S-QSN- 38
Db 312 AEVDAAGPAASHRYVLPGRYHVTAVALGAGSALLGTDVQVEAAPAALELVCPSSVQSD 371
QY 39 ----I-N--G-----AL-----V--P-----P----- 46
Db 372 SLDSLQNRGGSGLEAAYSIVALGEPARAVHPLCPSDTEIFPGNGHCYRLVVEKAAWLQ 431
QY 47 ----AA-----LIS-----I-----I-O--K-----G-----L-- 57
Db 432 AOEQOQAWAGAALAMWDSPAVQRFVLSRVTRSLDVWIGFSTVQGVVEVPAQGEAFSL 491
QY 58 -Q-----YV-E-----AE--V--S-- 65
Db 492 CQNLWFGPEHPATAEHCVRGLGPTGMCNTDLCAPHSYVCELOPGGVPQDAENLLVGAP 551
QY 66 ----IN--BDG-----TL--F----- 73
Db 552 DLQGLTPLAQODGLSAPHEPVVVFPGRLRLSREAFLTAEFGTQELRRPAQLRQVYR 611
QY 74 ----D-----GR-P-I----- 78
Db 612 LLSTAGTPENGSEPSRSPDNRTQLAPACMPGGRWCPGANICLPDASCHPOACANGCTS 671
QY 79 ----E--S-----LSLI-----DAV-M-P-DVV----- 92
Db 672 GPGCPAYALWREFLSVAAGPPAQYS-VTLHGQD-VLMLPQDVLVGLQHDAGPGALLHC 729
QY 93 ----Q--T-----R-----Q----- 96
Db 730 SPAPGHGPQAPYLSANASSWLPPLPAQLEGTWACPCALRLIAATEQLTVLLGLRPN 789
QY 97 ----QA-----Y-----RD-KL-----A--Q- 105
Db 790 LRMPGRYEVRAEYGVNGVSRHNLSCSPDVVSPVAGLRYVPAPRDRGLYYPTNGSALV 849
QY 106 ---- 107
Db 850 DSGANATATARWPGGSVARFENVCPALVATFVPGCPWETNDTLFSVVALPWLSEGHV 909
QY 108 ----A--A-----A-----A----- 113
Db 910 DVVENSASRANLSLRVTAEEPICGLRATPSPEARVLQGVLYRSPVVEAGSDMVFRTI 969
QY 114 ----AA-----A-----AT--N--O--Q----- 121
Db 970 NDQSLTFQNVFNVYQSAAVFKLSLTASNHVSNVTNVNVTVERNNRQGLQVSTVPA 1029
QY 122 ----G--SA-K--N--GE-----N-----TA----- 131
Db 1030 VLSFNATLVLTGGVLYDSAVEVAFLWNFGDGEALHQFQPPYNEFPVPDPSVAQVLVEH 1089
QY 132 N-----GE-----EN-----G-----A-----H 139
Db 1090 NVMHTVAAPCEYLLTVLASNAFENLAQOVFVSVRASLPSVAVGSDGLVAGRPVTFYPH 1149
QY 140 ----T-----I-----ANNHT----- 146

Db 1150 PLSPGGLVLTWDFGDSFVLTSQSPAANHYSASRGTYHVRLEVNNVTSGAAQAADRVVF 1209
QY 147 -----DM-M-E-----V-----D-GD----- 154
Db 1210 BELGLSDMSLAVEGAPVWVSAAVQTGDNITWTDFMDGDTVLSGPEATVEHVYLAQN 1269
QY 155 -----VE-----IPS-----NKA----- 162
Db 1270 CTVTGAASPAGHARSRLVLFVLEVRVEPAACIPTQPDARLTAYVTGNPAHYLFDWT 1329
QY 163 -----V-----L-R-G-----H-----ESEV----- 172
Db 1330 FGDGSNTTVRGCTVTHNFTSRGTPLALVLSRVNRAHYFTSICVEPEVGNVLQPER 1389
QY 173 -FI-----CAW-----NP----- 179
Db 1390 QFVLGDEALVACAPPPPYRYTDFGTGEAAAPTRAGPEVTFIYRDPGSYLVTVTASN 1449
QY 180 -VS---D-L-LV-S---GS-----GD----- 190
Db 1450 NISAANDSALVEQEPVLVTSIKVNGSLGLELQOPLYFSAGVGRGPASVYLDLGDGWL 1509
QY 191 -----S-----TARI-WN-L-SE---NST---SG-----P----- 206
Db 1510 GPEVTHAYNSTGDTVRVAGMNEVSRSEAWLNTYVRRVGLVYNASRTVVPNGSVSFS 1569
QY 207 T-----QL--VL--RH-C-----I--RE-G-----G--QD----- 221
Db 1570 TSLEAGSDVRYSWLDR-CTPIPGGPTISYTRFSVGTNIIIVTAENEVGSQDSIFVVV 1628
QY 222 -----V-----PSN-----XD---V---T-----SL----- 231
Db 1629 LQLEGLVQVGGGRYFPTNHTVQLQAVRDGTNVSYSWTAWDRGPALAGSGKGFSLTVL 1688
QY 232 -----D-----W-----N-----S--- 235
Db 1689 EAGTYHVQLRATNMLGSADCTMDFEVPVGLMVTASNPAAVNTSVTLSABELAGSGV 1748
QY 236 -----EG-----T-L-LAT-----GS-----YD----- 246
Db 1749 VYTWSLEGLSWETSEPTTHSTPGLHLVTVTGNPLGSANATVEVDVQVPVSGLSIR 1808
QY 247 -----G-F-A-----R-----I---W-----TK---D-G-----NL 258
Db 1809 ASBPGGSFVAAGSSVPFGQLATGTNVMCWAVPGGSSKRGPHVTWVPDAGTFSIRLN- 1867
QY 259 AS-----T-----L-----L-----GQ-H-----K-----G--- 267
Db 1868 ASNAVSWSATYNLTABEPIVGLVLWASSKVAVPGLVHFQILLAAAGSAVTFRLOVGGAN 1927
QY 268 P-I-----F--AL-----K--W-----N--KG----- 278
Db 1928 PEVLPGFRFSHPRVGDHVVSVRGKHNVSQAQVRIIVLEAVSGLQMPNCPCEGIATG 1987
QY 279 --NF-----ILS-----AG----- 285
Db 1988 TERNFTARVQGRSVAWAYFSYKQVQDSVLISGRDVTYTPVAAGLLEIQVRAFNALG 2047
QY 286 -----V-D-----K--T-----T-----II--WD----- 294
Db 2048 SENRTLVEQDAVQVVALQSGPCFTNRSQAPEAATSPSPRVAYHWDGDSFGQDQDTE 2107
QY 295 --A-HT--G--E--A-----KO--O-----FP----- 305
Db 2108 PRAEHSYLRPGDYRVQVNASNLVSFFVAQATVTVQLACKREPDVVPLQVLMRRSQRN 2167
QY 306 F--H-----SA-----P-----AL---DV----- 314
Db 2168 YLEAHVDLRDCTVYQTYRWEVYRTASCQRPGRPARVALPGVDVSRPRLVPLALPVGH 2227
QY 315 -----D-----W-----QS---N--- 319
Db 2228 YCFVFWVSGDTPLTQSIQANVTVAPELRLVPIEGGSYRWSDTRDLVLDGSSYDPNLE 2287
QY 320 -----N-----TF----- 322
Db 2288 DGDQTPLSFHWACVASTQREAGGCALNPGRGSSVTIIPRERLAAGVEYTFSLTVWKAGR 2347
QY 323 --A-----SC----- 325
Db 2348 KEEATNQTVLIRSGRPVTVLSCEVSCKAQAVYEVSRSSVYVLEGRCLNCSGSKRGWAA 2407
QY 326 --S--T--D-----M-----C--I----- 331
Db 2408 RTFSNKTLLVDTTTSTGSAGMRLVLRGVLRDGEGYTFTLTVLGRSGEEGCASIRLSP 2467
QY 332 -----H-----V-----C-- 334
Db 2468 NRPLGSGCRLPPLGAVHALTTKVHFECTGWDHDAEDAGAPLVVALLRRCRQGHCEFCV 2527
QY 335 -K-----L-----G-----QD-----R-----P-----I-- 342
Db 2528 YKGSLSYGAFLPPGFRPHFEVGLAVVQDQLGAAVVALNRSLAITLPEPNGSATGLTVM 2587
QY 343 --KT--FQG-----H-----T--NE--V-----N-- 353
Db 2588 LHGLTASVLPGLLRQADPOHVEISLALVTVLNEYERALDVAAEPKHERQRAQIRKNIT 2647
QY 354 -----WD-----PT--G-----NL-----A---I---K----- 356
Db 2648 ETLVSLRVHTVDDIQIIRAAALAQCMGPSRELVCRSCLKQTLHKLKLEAMMLILQAEATTAGTV 2707
QY 357 -----LA-----S-----C-----SD----- 369
Db 2708 TPTAIGDSILNITGDLIHLASSDVRAPQPSSELGABSPSRMVASQAYNLTSALMRILMRSR 2767
QY 364 -----LA-----S-----C-----SD----- 369
Db 2768 VLNEBPLTLAGEEIVAGKRSDPRLLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIF 2827
QY 370 --D-----D-----M--T-----L-----KI-----W----- 376
Db 2828 LVDSNPPFPGYISNYTVSTKVASMAFQIQAGAIPIERLASERAITVKVPNNSDMAARGH 2887
QY 377 -----SM-----KQ-----D----- 381
Db 2888 RSSANSANVVVQPOASVGAVVTLDSNPAAGLHLQLNLTLLDGHYLSBEPEPYLAVYLH 2947
QY 382 -----NC-----V-----H-----DLQA--H-N-----K---EI 394
Db 2948 SEPRNEHNCASRRIRPESLQADHRPYTFIFSGSRD-PAGSYHLNLSHFRNSALQV 3006
QY 395 -----YT-----IKW-----SP-----T--G----- 403
Db 3007 SVGLYTSLCOYFSEEDMV-WRTEGLLPLEETSPRQAVCLTRHLTAFASLFPVPPSHRVFV 3065
QY 404 --P----- 404
Db 3066 FPEPTADVNIYVMLTCAVCLVTVMMAAILHLKLDLDASGRRAIPFGQGRFKYBILVK 3125
QY 405 -----CT-----N----- 407
Db 3126 TQWGRSGGTAHVGMVYGVDSRSGHRHLDGDRAFHRNSLDIFRIATPHSLGSLWKIRVW 3185
QY 408 --N---P-----N-----AN--L--M--L--AS-AS--- 419
Db 3186 HDNKGSLSPANFLOHVIRDLQARSAPFLVNDWLSVETEANGGLVEKEVLAASDAALLRF 3245
QY 420 -----FD-----S--T--V--R-----L-----W-----D-- 428
Db 3246 RBLVAELQRFKDHIMLSIWDPRPSRFRTRIQATCCVLLICLFLGANAVMYGAVGDS 3305
QY 429 -----VD-----R-----G----- 432
Db 3306 AYSTGHVSRLSPLSVDTVAVGLVSSVVVYVYLAILFLFRMRSKVGSPPTPAGQOVL 3365

```
QY 433 -I--C--I--H-----TLT-----439
Db 3366 DIDSCLDSSVLDSSFLTFSGHAEAFVGQKSDLFLDDSKSLVCWPSGEGTLLSWPDLSD 3425
QY 440 -----K-----440
Db 3426 PSIVGNLRQLARGQAGHGLGPEEDGFSLASPSYSPAKSPFSADESDELIQQVLARGVSSPAP 3485
QY 441 -----H-----Q-----E---443
Db 3486 TQDTHMETDLLSSSTPGKETTALQRLGELGPPSPGLNWEQPOAARLSRTGLVEGLR 3545
QY 444 -----P-----V--Y-----SVA-----449
Db 3546 KRLLPAWCAISLAHGLSLLVAVAVAGVWGASFPFGVSVALLSSASFLASFLGWEPL 3605
QY 450 -----FG-----PD-----453
Db 3606 KVLEALYFSLVAKRLHPDEDDTLVESPAVTPVSARVVRPPHGFALFLAKEEARKVKR 3665
QY 454 -G-R--Y-----LAS-----G-----S-F-----462
Db 3666 LHGMLRSLVYMLFLVTLILASYGDASCHGHAYRLQSAIKQELHSRAFLAITRSEBELWPW 3725
QY 463 -----TO-----T-----G-----AL-----476
Db 3726 MAHVLLPYVHGNOSSPELGPRLRQVRLQALYDPPGPRVHTCSAAGGPFSDYDVGWE 3785
QY 467 --H-----I-----W--N--470
Db 3786 SPHNGSTWAYSAPDLLGAWSGSCAVYDGGVQELGSLSEBSRDLRLPLQLHNLNDR 3845
QY 471 -----TO-----T-----G-----AL-----476
Db 3846 SRAVLELTRYPAGVGHAAVTLRLFPAPAGRALAALSVPFALRRLSAGLSPLITSVC 3905
QY 477 -----VH-----SY-----R-----G-----T--483
Db 3906 LLLFAVFAVAEARTWHRGRWRVRLGAWARMLLVALTATALVRLAQLGAADROWTRF 3965
QY 484 --G-----G-----G-----IF-----E-488
Db 3966 VRGRPRFTSDQVAVSSAAGLAASLLFLLLVKAQHVRVFRQWSVFGKTLRALPEL 4025
QY 489 -----V-----V-----489
Db 4026 LGVTGLVLGVAYQAAILLVSSCVDSLWSVAQALLVLCPTGLTCLTCLPAESWHLSPLL 4085
QY 490 C-----WNA-----A--G-----D-----496
Db 4086 CVGLWALRLMGLNLGAVILRWYHALRGELYRPAWEPQDYEMVELFLRLRLMGLSKV 4145
QY 497 -----KV-----G-----A-S-AS-----DG-SV-----507
Db 4146 KEFRHKVRPEGMELPSRSGSKVSPDVPFPSPAGSDASHPTSSSQDLGLSVSLGRIGT 4205
QY 508 -C-----V-----L-----D-----L--RK 514
Db 4206 RCEPEPSRLOAVFEALITQFRLNQATEDVYQLEQLHSLQGR 4249
```

RESULT 15

US-09-052-469-6

; Sequence 6, Application US/09052469

; Patent No. 6380360

; GENERAL INFORMATION:

; APPLICANT: Harris et al.

; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE

; TITLE OF INVENTION: AND USES THEREOF

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Banner & Witcoff, Ltd.

; STREET: One Financial Center

```
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,469
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-April-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-052-469-6
```

Query Match 69.5%; Score 2548.4; DB 3; Length 4339;
Best Local Similarity 9.8%; pred. No. 1.1e-31;
Matches 412; Conservative 83; Mismatches 13; Indels 3676; Gaps 357;

```
QY 1 MSIS-----S-----D-----EVN-----F---10
Db 29 LDVSHNLLRALDVGLLANLSALAEALDISNNKISTLEEGIFANLNLSEINLSGNPFECDC 88
QY 11 -LVY--RY-----L--Q-----E-----SG-20
Db 89 GLAWLPRWABEQRVVVQPEAATCATGPGSLAGQPLLIGIPLLDGGCGEYVACLPDSSGT 148
QY 21 -----PS--H-----SAF-----E-----26
Db 149 VAAVSFAAHEGLQPEACSAFCFTGQGLAALSEQWCLCGAAQPSASFACLSLCSGP 208
QY 27 -----T-----F-----G-----I-----30
Db 209 PPPAPTCTRGPTLLQHVFPASPATLVGPHGLASQGLAAFHAPLPVTATRWDFGDGS 268
QY 31 -E-----SH-----I---S-QSN-38
Db 269 AEVDAAGPAASHRYVLPGRHYVTAVIALGAGSALLGTVDQVEAAPAALELVCFSSVQSD 328
QY 39 -----I-N-----G-----AL-----V-----P-----46
Db 329 SLDLSIQNRGGSGLEAAYSIVALGEBPARAVHPLCFSDTEIPFGNGHCYRLVVEKAALWQ 388
QY 47 -----AA-----LIS-----I---IQ--K-----G---L--57
Db 389 AQBQCAWAGAALAMVDSFAVQRFVLSRVTRSLDVMWIGFVSTVGVEVGPAPQGEAFSLES 448
```

QY 58 Q-----YV-E-----AE-V--S- 65
Db 449 CQNLPCPEHPATAEHCVRGLPTGWCNTDLCSPHYSVCELOPGQFVQDAENLLVCAPSG 508
QY 66 ---IN---EDG-----TL--P----- 73
Db 509 DLQGLTLPLAQDGLSAPHPEVVMVFPGLRLSREAFLLTAEFQTQELRRPAQLRLQVVR 568
QY 74 -----D-----CR--P--I----- 78
Db 569 LLSTAGTPENGSEPSRPNRTQLAPACMPGRCWCPGANICLPLDASCHPOCANGCTS 628
QY 79 -----E--S-----LSLI-----DAV-M-P-DV- 92
Db 629 GPGLPGAPYALWREPLFSVAAGPPAQYS-VTLHGQD-VLMLPGDVLGLQHDAGPGALLHC 686
QY 93 -----Q-----T-----R-----Q----- 96
Db 687 SPAPGHGPOAPYLSANASSWPLPPLPAQLEGTWACPCALRLLLAATEQLTVLLGLRPNG 746
QY 97 -----QA-----Y-----RD-KL-----A--Q- 105
Db 747 LRMPGRYEVRAEVNGVSRHNLSCSFDVVSVPAGLRVVPAPRDRGLYPTNGSALVLQV 806
QY 106 -----AA-----AT-----N-Q-Q-----OH-- 107
Db 807 DSGANATATARWPGGSVARSFENVCALVATFVPGCPWETNDTLFSVVALPMLSEGEHV 866
QY 108 -----A--A-----A-----A----- 113
Db 867 DVVENSASRANLSLRTAEPIICGLRATPSPPEARVLQGLVLRYSVPVBEAGSDMVFRTI 926
QY 114 -----AA-----A-----AT-----N-Q-Q----- 121
Db 927 NDQSLTFQNVFNVIYQGAAPKLSLTASNHSVNTVNYNTVERMRMQGLQVSTVPA 986
QY 122 -----G--SA-K--N--GE-----N--TA----- 131
Db 987 VLSPNATLVLTGVLVDSAVEAFLNFGDGEALHQFPNPNESFPVDPDPSVAQVLVEH 1046
QY 132 N-----GE-----EN-----G-----A-----H 139
Db 1047 NVMHTYAAPCEYLLTVLASNAFENLTQQVPVSVRASLPSVAVGVSDGLVAGRPTVFPYH 1106
QY 140 -----T-----I-----ANNHT----- 146
Db 1107 PLPSPGGVLYTWDFDGSFVLTSQSPAANHYSRGTYHVRLEVNNTVSGAAQAQDVRF 1166
QY 147 -----DM-M--E-----V-----D-GD----- 154
Db 1167 EELRGLSVDMSLAVEQAPVVVSAAVQTGDNITWTDFMDGDTGLSGPEATVEHVYLRQ 1226
QY 155 -----VE-----IPS-----NKA----- 162
Db 1227 CTVTGGAASPAGHARLSHLVLPVLEVRPEACIPTQDARLTAYVTGNPAHYLFDMWT 1286
QY 163 -----V-----V--L-R-G-----H-----ESEV----- 172
Db 1287 FGDGSSNTTVRGCTPTVTHNFTSRGTPLALVLSRVNRAHYFTSICVEPEVGNVTLOPER 1346
QY 173 -FI-----CAW-----NP----- 179
Db 1347 QFVLGDEAWLVACAMPPPPRYTWDFGTGEEAAPTRARGPEVTFIYRDFGSYLVTVTASN 1406
QY 180 -VS--D--L-----LV-S--GS-----GD----- 190
Db 1407 NISAANDSALVEQEPVLVTSIKVNSGLGLELQPPYLFSAVGRGRPASLYLDLGDGWNLE 1466
QY 191 -----S-----TARI--WN-L--SE--NST-----SG-----P----- 206
Db 1467 GPEVTHAYNSTGDTFVRVAGWNEVSRSEAWNLVTVKRRVRGLVNVNASRTVWPLNGSVSFS 1526

QY 207 T-----QL--VL--RHC-----I-----RE-G-----G--QD----- 221
Db 1527 TSLEAGSDVRYSWLVCDR-CTPIPGGPTISYFRSVGTNIIIVTAENEVGSAQDSIFVVV 1585
QY 222 -----V-----PSN-----KD--V--T-----S- 231
Db 1586 LQLIIEGLQVVGGRYFPPTNHTVQLQAVVRDGTNVSWTAWDRGPALAGSGKGFSLTVL 1645
QY 232 -----D-----D-----W-----N-----S-- 235
Db 1646 EAGTYHVQLRATNMLGSAWADCTMDPVEPVGMVMTTASNPAAVNTSVTLASLAGSGV 1705
QY 236 -----EG-----T--L--LAT-----GS-----YD----- 246
Db 1706 VYTWISLEGLSWETSEPTTHSPFPGLHLVMTAGNPLGSANATVEVDVQVPVSGLSIR 1765
QY 247 -----G-P--A-----R-----I--W-----TK--D-G-----NL 258
Db 1766 ASEPGGSFVAAGSSVFPFMGQLATGTNVSWCMWAVPGGSSKRGPHVTMVPFDAGTFSIRLN- 1824
QY 259 AS-----T-----L-----L-----GQ--H-----K-----G-- 267
Db 1825 ASNAVSWSATNLTABEPIVGLVWASSKVAVPQOLVHFQILLAAAGSAVTFRLQVGGAN 1884
QY 268 P-I-----F--AL-----K--W-----N--KKG----- 278
Db 1885 PEVLPGPFESHFPVRGDHVVSVRGKHNYSMAQAQVRIVVLEAVSGLQMPNCCPEGIATG 1944
QY 279 ---NF-----ILS-----AG----- 285
Db 1945 TERNFTARVQGRSVAYAWYFSLQVQGSVLSDVLTGRDVTYTPVAAGLLEIOVRAFNALG 2004
QY 286 -----V-D-----K--T-----T-----I--WD----- 294
Db 2005 SNRTLVLVEQAVQVVALQSGPCFTNRSQAPEATSPSPRRVAYHWDGSGSPGQDTE 2064
QY 295 --A-HT--G--E--A-----KQ--Q-----FP----- 305
Db 2065 PRAEHSYLRPGDYRVQVNASNLVSFFVAQATVTQVLACREPEVDVVLQVLMRSQRN 2124
QY 306 F--H-----SA-----P--AL--DV----- 314
Db 2125 YLEAHVLDLDCVTYQTEYRWEVYRTASQCPGRPARVALPGVDVSRPLVLPRLALPVGH 2184
QY 315 -----D-----D-----W-----QS--N-- 319
Db 2185 YCFVFWSPGDTPLTQSIQANVTVAPERLVPPIEGGSYRVWSDTRDLVLDGSESYDNLE 2244
QY 320 -----N-----N-----TF----- 322
Db 2245 DGDQTPLSFWACVASTQREAGGCALNPGRGSSVTIIPRERLAAGVEYTFSLTVWKAGR 2304
QY 323 --A-----SC----- 325
Db 2305 KEEATNQTILIRSGRPVIVSLEVCSCQAQAVYEVSRSSVYVLEGRCLNCSGSKRGWAA 2364
QY 326 ---S--T--D-----M-----C--I----- 331
Db 2365 RTFSNKTLLVDETTTSGSAGMRLVLRGVLRDGBGYTFTLVLRGSRBEEGCASIRLSP 2424
QY 332 -----H-----H-----V-----C-- 334
Db 2425 NRPPILGSGCRPLPLGNVHALTTKVHPECTGMDAEDAGAPLVYALLLRRRCROHCFCV 2484
QY 335 -K--L-----L--G-----QD-----R-----P-----I-- 342
Db 2485 YKGSLSYGAVLPPGFRPHFVEGLAVVQDQLGAAVVALNRSLAITLPBPNGSATGLTW 2544
QY 343 ---KT---PQ-----H-----T--NE-----V-----N-- 353
Db 2545 LHGLTASVLPGLLRQADPOHVIYSALVTLVNEVERALDVAEPKHQHQRAQIRKNIT 2604
QY 354 -----A-----I--K----- 356

Db 2605 ETLVSLRVHTVDDTQIIAALACQMGPSRELVCBSCLKQTLHKLKLEAMMLILQAEATTAGTV 2664
Qy 357 -----WD-----PT-G-----NL----- 363
Db 2665 TPTAIGDSILNITGDLIHASSDVRAPOPSLGAESPMSVQAYNLTSALMRLMSR 2724
Qy 364 -----LA-----S-----C-----SD----- 369
Db 2725 VLNEEPLTAGEEIVAOGRSDPSRLCYGAPGPGCHFSIPEAFSGALANLSDVOLIF 2784
Qy 370 -----D-----M-----T-----L-----KI-----W----- 376
Db 2785 LVDNPPFPYISNYTVSTKVASNAFQOAGAIPIERLASERAITVKVPNNSDWAARGH 2844
Qy 377 -----SM-----KQ-----D----- 381
Db 2845 RSSANSANVVVQFQASVAVVTLSSNPAAGHLQNLNYTLIDGHYLSERPYPYLAVLH 2904
Qy 382 -----NC-----V-----H-----DLQA-----H-N-----K-----EI 394
Db 2905 SEPRNEHCASRRIRPESLOGADHRPYTFIFSGSRD-PAGSYHNLNLSHFWSALQV 2963
Qy 395 -----YT-----IKW-----SP-----T-G----- 403
Db 2964 SVGLYTSLCQYFSEEDMV-WRTEGLLPLEETSPQAVCLTRHLTAFCASLFVPPSHVRFV 3022
Qy 404 -----P----- 404
Db 3023 FPEPTADVNYIVMLTCAVCLVYVMMAAILKLDQDASRAIPFCQGRKFYEILVK 3082
Qy 405 -----GT-----N----- 407
Db 3083 TGWRGSGTTHAGIMLYGVDSRCHRHLDGDRAFHRNSLDIFRIATPHSLGSLVKIRW 3142
Qy 408 -----N-----P-----N-----AN-L-M-L-AS-AS----- 419
Db 3143 HDNKGSLPAWFLQHVIVRDLQARSFAFLVNDMLSVETANGGLVEKEVLAASDAALLRF 3202
Qy 420 -----FD-----S-T-V-R-----L-W-----D- 428
Db 3203 RRLVAVELQRFKHIWLSIWRDPPRSRFRTRIQATCCVLLICLFLGANAVVYGAVDG 3262
Qy 429 -----VD-----R-----G----- 432
Db 3263 AYSGHVSRSLPSLSDTVAVGLVSVVVPYVLAIFLFRMSKVKVAGSPSPAGQOVL 3322
Qy 433 -----I-C-----I-----H-----TLT----- 439
Db 3323 DIDSCLDSSVLDSFLTFSGLHAEAFVGMKSDLFLDSDSKLVCWPSGEGTSLWPDLSD 3382
Qy 440 -----K----- 440
Db 3383 PSIVGSNRLQARQAGHLGPEDGFSLASPSYSPAKSFSASDEDLIQVLAEGVSSAP 3442
Qy 441 -----H-----O-----E----- 443
Db 3443 TQDTHMETDLSSLSSTPGEKTELTALQRLGELGPPSPGLNWEQQAARLSRTGLVEGLR 3502
Qy 444 -----P-----V-Y-----SVA----- 449
Db 3503 KRLLPAWCASLAHGLSLLVAVAVAGVGVGASFPFGVSVALLSSSASFLASFLGWEPL 3562
Qy 450 -----FS-----PD----- 453
Db 3563 KVLEALYFSLVAKRLHPDSDDTLVESPAVTPVSARVVRPVRPHGFALFLAKEEARKVR 3622
Qy 454 -----G-R-----Y-----LAS-----G-----S-F----- 462
Db 3623 LHGMRLSLVYMLFLLVTLASYGDASCHGHAYRLQSAIKQELHSRAFLAITRSEELWPW 3682
Qy 463 -----DK-----C-----V----- 466

Db 3683 MAHVLLPYVHGNQSSPELGPPRLRQVRLOEALYPDPGPRVHTCSAAGFSTSDYDVGWE 3742
Qy 467 -----H-----I-----W-N----- 470
Db 3743 SPHNGSGTWAYSAPDLLGAWMGSCAVYDSGGYVQELGLSLEBSRDLRFLQLHNLNDR 3802
Qy 471 -----TQ-----T-----G-----AL----- 476
Db 3803 SRAVLELTRYSPAVGLHAAVTLRLEFFPAAGRALAALSVPFALRRLSAGLSLPLTSVC 3862
Qy 477 -----VH-----SY-----R-----G-----T----- 483
Db 3863 LLLFAVFAVAEARTWHREGRWVLELGAWARWLLVALTAAATVLAQLGAADROWTRF 3922
Qy 484 -----G-----G-----IF-----E- 488
Db 3923 VRGRPRFTSFDOVAHVSSAAGLAASLLFLLLVKAACHVRFVQWVSFGKTLCRALPEL 3982
Qy 489 -----V----- 489
Db 3983 LGVTGLVVLGVAYAOIAILLVSSCVDSLWSVAQALLVLCPGTGLSTLCPAESWHLSPLL 4042
Qy 490 -----WNA-----A-G-----D----- 496
Db 4043 CVGLMALRLWGLALGLGAVILRWYHALRGELYRPAWEPQDYEMVELFLRRLRLWGLSKV 4102
Qy 497 -----KV-----G-----A-S-AS-----DG-SV----- 507
Db 4103 KEFRHKVRPEGMEPLPSRSSGSKVSPDVPVPPSAGSDASHPSTSSQLDGLSVSLRLGT 4162
Qy 508 -----C-----V-----L-----D-----L-RK 514
Db 4163 RCEPEPSRLOAVFEALLTOFDRLNQATEDVYQLEQQLHSLOGRR 4206

Search completed: January 3, 2005, 15:30:42
Job time : 93.6667 secs

ALIGNMENTS

Db 172 ATAAATTSAGVSHQNPKNREATVNGENRAHSV--NNHA-KPMEIDGEVEIPSSKATVLR 229
QY 167 GHESEVIFCAWNPVSDLLVSGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNK 226
Db 230 GHESEVIFCAWNPVSDLLVSGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGHDFPSNK 289
QY 227 DVTSLDWSNGTLLATSGYDGFARIWTKDGNLSTLQHGKPIFALKWKNKGNFIISAGV 286
Db 290 DVTSLDWSNGTLLATSGYDGFARIWTKDGNLSTLQHGKPIFALKWKNKGNFIISAGV 349
QY 287 DKTTIIWDAHTGEAKQFPPHSPALPDVDMQ--TPASCS--TDMCIHVCKLGDORPIKTF 345
Db 350 DKTTIIWDAHTGEAKQFPPHSPALPDVDMQ--NTTTFASCS--TDMCIHVCKLGDORPIKTF 408
QY 346 QGHTNEVNAIKWDPGNTLLASCSDDMTLTKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPG 405
Db 409 QGHTNEVNAIKWDPGNTLLASCSDDMTLTKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPA 468
QY 406 TNNPNANMLASASFDSTVRLWDVIRGICHTLTKHQPVSVAFPSDGRYLAGSFDKC 465
Db 469 TSNPNSNIMLASASFDSTVRLWDVIRGICHTLTKHQPVSVAFPSDGRYLAGSFDKC 528
QY 466 VHIWNTOTGALVHSYRGTGIFEVCMNAAGDKVGSASDGSVCVLDLRK 514
Db 529 VHIWNTOTGALVHSYRGTGIFEVCMNAAGDKVGSASDGSVCVLDLRK 577

RESULT 2

US-10-264-049-2892
; Sequence 2892, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2892
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (45)_
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2892

Query Match 88.9%; Score 3260.6; DB 15; Length 542;
Best Local Similarity 80.7%; Pred. No. 5.7e-47;
Matches 442; Conservative 40; Mismatches 22; Indels 44; Gaps 15;

QY 5 SDEVNF---LV-----YR-----YLOESGFSAFTFGIESHISQSNNG 41
Db 1 TDC--FKESKVDLQSQKPLRKTPTLTCGFCGTGLTC--GFSAFTFGIESHISQSNNG 57
QY 42 ALVPPAALISIIQKGLQYVBAEVSINEDGTLFGRIPIESLSLIDAVMPDVVTRQAYRD 101
Db 58 TLVPPAALISIIQKGLQYVBAEVSINEDGTVFDPRIPIESLSLIDAVMPDVVTRQAFRE 117
QY 102 KLAQOH--AAAAAAA-----AATNO--QG--S-----AKNGENTANGENGAHTTANNHTDM 148
Db 118 KLAQQQASAAAAAATAATAATTSAGVSHQNPKNREATVNGENRAHSV--NNHA-K 175
QY 149 -MEVDGVEIIPSKAVVLRGHESEVFTCAWNPVSDLLVSGSGDSTARIWNLSNSTSGPT 207
Db 176 PMEIDGEVEIPSSKATVLRGHESEVFTCAWNPVSDLLVSGSGDSTARIWNLSNSTSGPT 235

QY 208 QLVLRHCIREGGQDVPSNKDVTSLDWSNGTLLATSGYDGFARIWTKDGNLSTLQHGK 267
Db 236 QLVLRHCIREGGHDFPSNKDVTSLDWSNGTLLATSGYDGFARIWTKDGNLSTLQHGK 295
QY 268 PIPALKWKNKGNFIISAGVDKTTIIWDAHTGEAKQFPPHSPALPDVDMQ--TPASCS 326
Db 296 PIPALKWKNKGNFIISAGVDKTTIIWDAHTGEAKQFPPHSPALPDVDMQ--NTTTFASCS 354
QY 327 TDMCIHVCKLGDORPIKTFQGHTEVNAIKWDPGNTLLASCSDDMTLTKIWSMKQDNCVH 386
Db 355 TDMCIHVCKLGDORPIKTFQGHTEVNAIKWDPGNTLLASCSDDMTLTKIWSMKQDNCVH 414
QY 387 LOAHNKEIYTIKWSPTGNTNNPNANMLASASFDSTVRLWDVIRGICHTLTKHQPVS 446
Db 415 LOAHNKEIYTIKWSPTGNTNNPNANMLASASFDSTVRLWDVIRGICHTLTKHQPVS 474
QY 447 SVAFSPDGRYLAGSFDKCVHIWNTOTGALVHSYRGTGIFEVCMNAAGDKVGSASDGS 506
Db 475 SVAFSPDGRYLAGSFDKCVHIWNTOTGALVHSYRGTGIFEVCMNAAGDKVGSASDGS 534
QY 507 VCVLDLRK 514
Db 535 VCVLDLRK 542

RESULT 3

US-10-363-829-446
; Sequence 446, Application US/10363829
; Publication No. US20040142331A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;
; APPLICANT: Altus, Christina M.; Dufour, Gerard E.;
; APPLICANT: Chalup, Michael S.; Jackson, Jennifer L.;
; APPLICANT: Jones, Anissa L.; Yu, Jimmy Y.;
; APPLICANT: Wright, Rachel J.; Gietzen, Darryl;
; APPLICANT: Liu, Tommy P.; Yap, Pierre E.;
; APPLICANT: Dahl, Christopher R.; Momiyama, Monika G.;
; APPLICANT: Bradley, Diana L.; Rohatgi, Sameer D.;
; APPLICANT: Harris, Bernard; Roseberry Lincoln, Ann M.;
; APPLICANT: Gerstin, Jr., Edward H.; Peralta, Careyna H.;
; APPLICANT: David, Marie H.; Panzer, Scott R.;
; APPLICANT: Flores, Vincent Z.; Daffo, Abel;
; APPLICANT: Marwaha, Rakesh; Chen, Alice J.;
; APPLICANT: Chang, Simon C.; Au, Alan P.;
; APPLICANT: Inman, Rebekah R.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PT-1183 USN
; CURRENT APPLICATION NUMBER: US/10/363,829
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: PCT/US01/27628
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/229,751
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,749
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,750
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,747
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,748
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,583
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,517
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,610
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,597
; PRIOR FILING DATE: 2000-09-06
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: PERL Program


```
; SEQ ID NO 446
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: L1:021759.1.orf2:2000SEP08
US-10-363-829-446

Query Match      74.0%; Score 2716; DB 16; Length 395;
Best Local Similarity 98.5%; Pred. No. 2.1e-37;
Matches 385; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSISDEVNVLVRYLQESGFSAFTFGIESHSOSNINGALVPPAALISIIQKGLQYV 60
DB 5 MSISDEVNVLVRYLQESGFSAFTFGIESHSOSNINGALVPPAALISIIQKGLQYV 64
QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQVQROAQYRDKLAQCHAAHAAAAAATNQ 120
DB 65 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQVQROAQYRDKLAQCHAAHAAAAAASQ 124
QY 121 QGSAKNGENTANGFCAHTIANNHTDMMEVDGVEIPSNKAVVLRGHSEVEFICAWN 180
DB 125 QGSAKNGENTANGFCAHTIANNHTDMMEVDGVEIPSNKAVVLRGHSEVEFICAWN 184
QY 181 SLLVSGSGDSTARIWNLSNSTSGTQVLVRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
DB 185 SLLVSGSGDSTARIWNLSNSTSGTQVLVRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244
QY 241 ATGSDGFARITWKGNGLASTLQGHKGPFPALKWKKGNFILSAGVDKTTIWDHAHTGEA 300
DB 245 ATGSDGFARITWKGNGLASTLQGHKGPFPALKWKKGNFILSAGVDKTTIWDHAHTGEA 304
QY 301 KOQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
DB 305 KOQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 364
QY 361 GNLLASCSDDMTLKINSMKQDNCVHDLQAHN 391
DB 365 GNLLASCSDDMTLKINSMKQDNCVHDLQAHN 395

RESULT 4
US-10-451-168-78
; Sequence 78 Application US/10451168
; Publication No. US20040091969A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GF50039
; CURRENT APPLICATION NUMBER: US/10/451,168
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: PCT/US01/49232
; PRIOR FILING DATE: 2000-12-17
; PRIOR APPLICATION NUMBER: 60/256,710
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/257,048
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/260,482
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/264,922
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/266,797
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/276,988
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/281,535
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/289,622
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 110

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 5635
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-451-168-78

Query Match      71.7%; Score 2631.9; DB 15; Length 5635;
Best Local Similarity 9.6%; Pred. No. 8.2e-30;
Matches 440; Conservative 57; Mismatches 15; Indels 4071; Gaps 372;

QY 1 M-S-----I-----SS---D-----EV----- 8
DB 1 MISWEVHTVFLFALLYSLAQDASQPSIRABEIEGASTLAFVFDVTGSMYDDLQVVI 60
QY 9 -----NF-LV----- 12
DB 61 EGASKILETSLKRPKRPLNFALVPHDPFPIGVTITDPKFKQYELRELYVGGGDCPE 120
QY 13 -----Y-----R---Y-----LQ----- 17
DB 121 MSIGAIIKIAIEISLPGSFYVFTDARSKDYRLTHEVLQIIQOKSQVVFVLTGDCDDRTH 180
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
DB 181 IGYKYVEETASTSSGQVPHLDKQVNEVLKWVEEAQASKVHLLSTDLHLEQAVNTWRIPF 240
QY 27 -----T-----I-----EG----- 31
DB 241 DPSLKEVTVSLSGSPSMIEIRNPLGLIKKFGHLHLLNIHNSAKVNVNVEPEAGMWTVK 300
QY 32 ---S--H-----I-----S---Q----- 36
DB 301 TSSSGHSHVRITGLSTIDFRAGFSRKPTLDFKTVSRPVQGIPTVYLLNTSGISTPARID 360
QY 37 -----S-----NI-----N-----G----- 41
DB 361 LLELLISGSSSLKTIPIVKYYPHRKPYGIWNISDFVPPNEAFFLKVGTGYDKDYLFORVSS 420
QY 42 -----A-----L---VP-----P-----A 47
DB 421 VFSSTIVDPAPKVTMPEKTPGYVYLPQGPSVDSLLPFTLSFVRNGVTGLGVQYLKESA 480
QY 48 -----A---L-----I-----S---IIQ-----KG----- 56
DB 481 SVNLDIAKVTLSDEGFYECIAVSSAGTGRAQTFDVSPPVQVNNVTVTGERAVLT 540
QY 57 -----L--Q-----Y---V--E----- 61
DB 541 CLIIISAVDYNLTWQRNDRVRLAEPARIRTLANLSLELKSVPKENDAGEYHCVSSEGGSS 600
QY 62 -A-----EVS-----NE----- 68
DB 601 AASVFLTVQEPKVTVMKPNQSFSGSEVSIMCSATGYPKPKIAWTNDFMFI VGSHRM 660
QY 69 --DGTLF-----D-----G-----R---P-----I-- 78
DB 661 TSDGTLFIKNAAPKADAGIYGCCLASNSAGTQKONSTLRYEAPKLMVVSQSELLVALGDI 720
QY 79 -E---S-----L--S--LI-----DA---V----- 87
DB 721 MECKTSGIPPPQVWFKGDLRLPSTFLIIDLGLLKIQTQDLQDAGDYTCVAINBAGR 780
QY 88 -----M-----P-DV-----VQ-----T----- 94
DB 781 ATGKITLDVGSPPVFIQEPADVSMIEGNSVTLPCYVQGYPEPTIKWRRLDNMPFISRPFS 840
QY 95 -----R----- 95
DB 841 VSSISQIRGALFILNLWASDKGTICEAENQFGKIQSETTVTGTGLVAPLIGISPSVAN 900
QY 96 ---QQ-----A-----Y---R-D-KL-----AQ----- 105
```

901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLLQNPYITVRSQSLHIERVQLQDGGBYTC 960
106 -----QH-----A----- 108
961 VASVAGTNNKTTSVVVHVLFTIQHQOILSTIEGIPVTLPCASGNPKPSVIMSKGEL 1020
109 -----A--A--A-----A--A--A----- 114
1021 ISTSSAKFAGADGSLVYVSPGGESGEYVCTATNAGYAKRKVOLTIVYVPRVFGDQ 1080
115 -----A-----A----- 116
1081 LSQDKPVEISVLAGEEVTLPCEVKSPLPPITWAKETQLISPFSPRHFTLPSPGSMKITET 1140
117 -----ATN-----Q-----Q-GSA--K-----N----- 126
1141 RTSDSGYLCVATNIAGNVQAVKLVHVPKIQRGPKHLKVOVGQVDPICPNAQGTPLP 1200
127 -----GE--N-----T--A--N--G----- 133
1201 VITWSKGGSTWLDGEHHVSNPDGTLSDQATPSDAGIYTCVATNAGTDETEILHVQE 1260
134 -----E--N-----GA-----H----- 139
1261 PPTVEDLEPPYNTTFQERVANQRIEPPCPAKTPKPTIKWLHNGRELTGREPGISILEDG 1320
140 T--IA-----N-----N--HT-----DM----- 148
1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKDEQEVNTVSVLLNQLTNL 1380
149 -----M-----E----- 150
1381 FCEVEGTPSPILIMYKDNQVTSSTIQVNNKILKLFRATPEDAGRYSCKAINAGTS 1440
151 -----VD-----G-----DV--E-----I-----P-- 158
1441 OKYFNIDVLVPTIIGTNFNEVSVLNRDVALEQVKGTPFPDIHWFKDGKFLFLGDPN 1500
159 -----SN--K--A-----V-----VL 165
1501 VELLDRGQVLHLKNARNDKRYQCTVSNAGQAKDKIKTIYNPPSIKGNVTDISVL 1560
166 -----RG----- 168
1561 INSLIKLECETRGLPWPAITWYKDGQPIIMSSQALYIDKQYLHIIPRAQVSDSATYTCV 1620
169 -----E-S-----E-----VFI-----C--A-----W----- 177
1621 ANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECKAAGNPSPILTWLKDG 1680
178 -----N----- 182
1681 PVKANDNIRIEAGKKLEINSAQEI DRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740
183 -----LL-----VSGS-----G--D-----S-- 191
1741 TSYFIVMNNLELDCVTSPPPTIMWLKDGQLIDBERDGFILLNGRKLVIQAQVSN 1800
192 -----TA-----R----- 196
1801 GLYRCMAANTAGDHKKEFEVTVHVPPTIKSSGLSERVVVVKYKPVALQCIANGIPNPSITW 1860
197 -----NL--S-----E-----NS--T-----S 204
1861 LKDDQPVNTAGNKLKIQSSGRVLQIAKTLLDAGRYTCVATNAGETQQHIQLHVHPPS 1920
205 -----G-----PTQL-----V-----L-R----- 212
1921 LEDAGKMLNETVLSNPVQLECKAAGNPVITWYKDNRLLSGTSMTFLNRQIIDIES 1980
213 -----H----- 215
1981 AQISDAGIYKCVAINSAGATELFYSLQHVAPISGNSNNMVVVVNNPVRLECEARGIPA 2040

216 -----R-----E--G-----G--Q--D--- 221
2041 PSUTWLKQDGPVSSFFNGLOVLSGGRIALATSAQISDTGRYTCVAVNAAGEKORDILRV 2100
222 -VPSN-----KD-----VT---S-L- 231
2101 YVPFNMIGBEQNVSVLISOAVELLQSDAIPPTTLTWLKDGHPLLPGLSISENISVLK 2160
232 -----D-----W--N--S-----EG--TLL--A 241
2161 IEDAQVQDGRYTCHEATNVAGTERKNYVNIWPPNIGGSDTLTQLTVEIGNLILICES 2220
242 T-----GS--Y-D-----G-----F-----A-- 249
2221 SGIPPNLIWKKGSPVLTDMSGRVRLSGGROLOISIAEKSDAALYSCVASNAVGTAKK 2280
250 -----R-----I-----I-----WTQD----- 256
2281 EYNLOVIRPTITNSGSHPTIIVTRGKSI SLECEVOGIPPTVTWMDGHPLIKAKGVE 2340
257 -----NL-AS-----T-----L-----GQH----- 265
2341 ILDEGHILQLKNIHVS DTRYCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPENISVVE 2400
266 K-----G--P--I--FALK--W----- 274
2401 KNSVLTCEASGIPLSITWPF--KOGWPFVLSNSVRILSGRMLRLMQTTMEDAGQYTCV 2458
275 -----N-----K-K-----GN----- 279
2459 VRNAGEERKIFGLSVLPVPHVIGENTLEDVKVEKQSVTLTCEVTGNPVEITWHDQ 2518
280 -----F--I-----L-----S--AGV-- 286
2519 PLODEAHHIISGGRFLQITNVQVPHTGRTYTCCLASSPAGKSRFSLSNFVSPITAGVGS 2578
287 D--K--T-----T--I-----I--W----- 293
2579 DGNPVDVTILNSPTSLVCEAYSPPATITWFKDGTPLESNRIRILPGGRTILOILNAQE 2638
294 D-----A-----H-----TG-----EA 300
2639 DNAGRYSCVATNEAGEMIKHYEVKVIPIPIINKDLWPGLSPEVKIKVNNLTLECEA 2698
301 -----K-----Q-----Q-----F 304
2699 YAIPSASLSWYKDGQPLKSDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELDF 2758
305 -----P--F----- 307
2759 DVNIQVPPSFQKLWEIGNMLDTCRNGEAKDVIINNPI SLYCETNAAPPPTLTWYKDGHL 2818
308 -SA-----P-----A-----LDV-----D- 315
2819 TSSDKVLLILPGRVLIQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLPPIKANSDL 2878
316 -----WQ-----SN----- 319
2879 PEEVTVLVNKSALIECLSSGSPAPRNSWQKQPLEDDHHKFLSNGRILQILNTQITDI 2938
320 -----NT-----PAS--C-----S- 326
2939 GRVYCVAEANTAGSAKKYFNLVHVPSPVIGPKSENITVVVNNFISLTCEVSGFPPLDSW 2998
327 -----T-----D-----MCI----- 331
2999 LKNEQPIKLTNTLIVPGGRTLIQIRAKVSDGGEYTCIAINOAGESKKKFSLTIVVYVPSI 3058
332 --H-----V--C-----K-----L--GO--- 338
3059 KOHDSLSLVNVNREGTSVSLCESNAPPPVITWYKNGRMITESTHVEILLADQOMLHIK 3118

QY 339 ---D---R---P-IK---TF--- 345
Db 3119 KAEVSDTGQYVCRAINVAGRDDKNFHLNVVPSIEGPEREVIVETISNPTVLTCDATGI 3178
QY 346 ---Q---G---H---T--- 349
Db 3179 PPPTIAWLKHKRIENSLEVRILSGSKLOIARSOHSDSGNYTCIASNMEGKAQYF 3238
QY 350 ---N-E-V-NA---IKW-D-P-T--- 360
Db 3239 LSIQVPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIOWLKDGKPIASGETERIRVS 3298
QY 361 ---G---NLL-A-S---C-S---D---MTL--- 373
Db 3299 ANGSTLIYCALATSDTKYTCVATNPAGEEDRIFNLNVVPTPIRGNKDEAKLMTLVD 3358
QY 374 ---K---I-W---SM--- 378
Db 3359 SINIECRATGPPQINWLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVVTCVASNR 3418
QY 379 ---K---Q---DN---CV--- 384
Db 3419 AGVDNKHYNLOVPAPPNMDNSMGTETITVLKGSSTMACITDGTAPAPMAWLEDGQPLGL 3478
QY 385 ---H---D---L---QA---H-N---KEI 394
Db 3479 DAHLTVSTHGMVLQQLKAETEDSGKYTCIASNEAGEVSKHFIKLVLEPPHINGSBEEH 3538
QY 395 ---YT---I---K-W--- 399
Db 3539 SVIVNNPLELTCIASGPAPKMTWMDGRPLPQTDQVOTLGGGEVLRISTAQVEDTGRYT 3598
QY 400 ---SPTG---P---GT---N--- 408
Db 3599 CLASSPAGDDDKBYLVRVHPVPIAGDEPRDITVLNRQVLTLECKSDAVPPVITWLRN 3658
QY 409 ---P---N---ANLM-LAS-A---S-F--- 420
Db 3659 GERLOATPRVRLISGGRYLQINNADLGDNTANYTCVASNIAGKTTREFILVNVPPNIKGG 3718
QY 421 ---DSTV---RL-W-D-V---DR---G-I--- 433
Db 3719 POSLVLINKSTVLECIAEGVPTPRITWRKDGAVLAGNHARYSILENGFLHIOAHVTD 3778
QY 434 ---C---I---H---TL---T--- 439
Db 3779 GRYLCMATNAGTDRRIDLOVHPVPSIAGPNTMTVIVNVQTLTACEATGIPKPSINWR 3838
QY 440 K-H---Q-E---P-V---Y---S 447
Db 3839 KNGHLLNVQONSRYLLSSGSLVIIISPSYDDTATYECTVTNGAGDDKRTVDLTVOVPPS 3898
QY 448 VA---F-S---P-D--- 453
Db 3899 IADEPTDFLVTHAPAVITCTASGVPPPSIHWTKNGIRLLPRGDGYRILSSGAIEILATQ 3958
QY 454 ---GRY---L---ASG-S-F 462
Db 3959 LNHAGRYTCVARNAGSAHRVTLHVHPPVIOQPSELVILNPIILLPCEATGTPSPF 4018
QY 463 ---D---KCV--- 466
Db 4019 ITWQKEGINVNTSGRNVHVLPSGGLQISRAVREDAGTYMCVAQNPGATLKGKILNVQVP 4078
QY 467 ---H---I-W--- 469
Db 4079 PVISPHLKEVIAVDKPIITLSCADGLPPDPDITWHDGRAIVESIRQVLSGSLQIAFV 4138
QY 470 ---N---T---Q--- 472
Db 4139 QPGDAGHYTCMAANVAGSSSTKLTAVHPPVIRIRSTEGHYTVNENSAIILPCVADGIPTP 4198
QY 473 ---T---G---A---L-VH- 478

Db 4199 AINWKDNVLLANLLGKYTAEPYGELILENVVLEDSGFYTCVANNAAGEDTHTVSLTVHV 4258
QY 479 ---S---Y---R- 481
Db 4259 LPTFTLPGDVSLNKEGEQLRLSCKATGIPLPKLTWTFNNNIIIPAHFDSVNGHSELVIERV 4318
QY 482 ---GT---G---G---IF---E---V-C--- 490
Db 4319 SKEDSGTYVCTAENSUGFVKAIGFVVVVPKPGDYPSNWIEPLGNGAILNCEVKGDP 4378
QY 491 ---WN---A---AGD--- 496
Db 4379 PTIQNNRKGVDIHSHIRQLGNSLAIYGTVNEADAGDYTCVATNEAGVVVERSMSTLQS 4438
QY 497 ---K-V---G---A---SAS--- 503
Db 4439 PPIITLEPVTINAGKILNCOATGEPQPTITWSRQSHISWDDRVNVLSNNSLYIAD 4498
QY 504 ---D---GSV---C-V--- 509
Db 4499 AQKEDTSEFECVARNLMSGVLVRVPVIVQVHGGFSQMSAWRACSVTCGKGIQKRSRLCNQ 4558
QY 510 -L-----DL-R-----K 514
Db 4559 PLPANGGKPCQSGDLEMRNCQNK 4581

RESULT 5

US-10-032-189-128
; Sequence 128, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard A
; APPLICANT: Grosse, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandes, Elma R
; APPLICANT: Rieger, Daniel K
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Millet, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenna
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826

; PRIOR FILING DATE: 2001-03-22
 ; PRIOR APPLICATION NUMBER: 60/279,840
 ; PRIOR FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: 60/282,981
 ; PRIOR FILING DATE: 2001-04-11
 ; PRIOR APPLICATION NUMBER: 60/283,656
 ; PRIOR FILING DATE: 2001-04-13
 ; PRIOR APPLICATION NUMBER: 60/309,247
 ; PRIOR FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: 60/311,754
 ; PRIOR FILING DATE: 2001-08-17
 ; PRIOR APPLICATION NUMBER: 60/313,331
 ; PRIOR FILING DATE: 2001-08-17
 ; NUMBER OF SEQ ID NOS: 260
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 128
 ; LENGTH: 5636
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (3003)
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (3041)
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (3367)
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.
 ; US-10-032-189-128

Query Match 71.6%; Score 2627.8; DB 14; Length 5636;
 Best Local Similarity 9.6%; Pred. No. 9.8e-30;
 Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;
 1 M-S-----I-SS--D--EV-----8
 1 MISVHVHTVFLFALLYSSLAQDASQSEIRAEFFEGASTLAFVDPVTGSMYDDLQVVI 60
 9 -----NF-LV-----12
 61 EGAKILETLKRPKRPLNFALVPHDPHDPBPGVTTITDPKQFQYELRELYVQGGDCPE 120
 13 -----Y-----R--Y--LQ-----17
 121 MSGAIKALEISLPGSFYVFTDARKSKYRLTHEVLQLIQKOSQVVFVLGTGCDRTH 180
 18 -----E--SG-F--S-H-S--A-----F 26
 181 IGVYVEEIASTSGQVFLDKKQVNEVLKWEAVQAKVHLLSTDHLEQAVNTWRIPF 240
 27 -----T-----FG--I-----E-----31
 241 DPSLKEVTVLSGSPMIEIRNPLGLIKKGFGHLHLLNHSKAVVNVKPEAGWTVK 300
 32 ---S-H-----I-----S--Q-----36
 301 TSSGRHSVRITGLSTIDFRAGFSRKPTLDPKTVSRPVGQIPYVLLNTSGISTPARID 360
 37 -----S-----NI-----N-----G-----41
 361 LLELLSISGSSSLKTI PVKYYPHRKPYGIWNI SDVPNEAFFLKVGTGYDKDYLFORVSS 420
 42 -----A-----L--VP-----P-----A 47
 421 VSFSSIVDPAPKVTMPKTPGYLQPGQIPCSVDSLPLPFTLSFVRNGVTLGVQYLKESA 480
 48 -----A--L-----I-----S--IIQ-----KG-----56
 481 SVSLDIATKVTLSDEGFECYAVSAGTGRAQTFFDVSEPPPVIVQVNNVTVPGERAVLT 540

QY 57 -----L-Q-----Y--V--E-- 61
 DB 541 CLIIISAVDYNLTWQRNDRVRLAEPARIRTLANLSLELKSVKFNDAGEYHCVSSSEGSS 600
 QY 62 -A-----EVS-I-----NE----- 68
 DB 601 AASVFLTVQEPKVTVMKPNQSF TGGSEVSI MCSATGYPKPKIAMTVNDMFIVGSHRYRM 660
 QY 69 --DGLTF-----D-----G-----R--P-----I-- 78
 DB 661 TSDGTLFIKNAAPKADAGIYGCCLASNSAGTDKQNSTURYIEAPKLMVQSELLVALGDITV 720
 QY 79 -E--S-----L-S-LI-----DA-----V----- 87
 DB 721 MECKTSGIPPPQVWKFGDLELRPSTFLIDPLGLLKLKIQTQDLADAGYTCVAINAEAGR 780
 QY 88 -----M-----P-DV-----VQ-----T----- 94
 DB 781 ATGKITLDVGSPPVFTIQEPADVSMIEIGSNVTLPCYVQGYPEPTIKWRRLDNMFIFSRPFS 840
 QY 95 -----R----- 95
 DB 841 VSSISQRLTGALFILNLWASDKGTIYCEABNOFGKIQSETTVTVTGLVAPLIGISPSVAN 900
 QY 96 -----QQ-----A-----Y--R-D-KL--AQ----- 105
 DB 901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLLQNPHYITVRS DGLHIERVQLQDGGEYTC 960
 QY 106 -----OH-----A----- 108
 DB 961 VASNVAGTNKTTVVVHVLTPTIQHGQILSTIEGIPVTLPCCKASGNPKPSVIWKKGEL 1020
 QY 109 -----A--A--A-----A--A--A----- 114
 DB 1021 ISTSSAKFSAGADGSLYVVSPEGESGEYVCTATNTAGYAKRKVQLTVYVPRVFGDLRG 1080
 QY 115 -----A-----A----- 116
 DB 1081 LSQDKPVEISVLAGEBVTLPCEVKSLPPPIITWAKETQLISPFSPRHTLPSPGSMKITET 1140
 QY 117 -----NTN-----Q-----O-GSA-K-----N----- 126
 DB 1141 RTSDSGMYLCVATNTIAGNTQAVKLVNVPVKIQRGPKHLKVQGVORVDPICNAQGTPLP 1200
 QY 127 -----GE--N-----T--A-N--G----- 133
 DB 1201 VITWSKGSMTMLVDGEHHVSNPDGTLSDIQA TPDSAGIYTCVATNTIAGTDETEILHVQE 1260
 QY 134 -----E--N-----GA-----H----- 139
 DB 1261 PPTVEDLEPPYNTTFOERVANQRIEFPCKAPKTPKTIKWLHNGRELTGREPGISILEEG 1320
 QY 140 T--IA-----N-----N--HT-----DM----- 148
 DB 1321 TLLVIASVTPYDNGEYICVAVNEAGTERKYNLKVHVPVPIKDKEQVSNVLLNQLTNL 1380
 QY 149 -----M-----SN--K-A-----V----- 150
 DB 1381 FCEVEGTSPSIIMWYKDNQVVTESSTIQTVNNGKILKLFRTATPEDAGRYSCKAINIAGTS 1440
 QY 151 -----VD-----G-----DV--E-----I-----P- 158
 DB 1441 QKYFNIDVLVPPTIIGTNFPEKVS VVLRDVALEQCVKGTPTTPIHFKDGPFLGLDPN 1500
 QY 159 -----RG-----SN--K-A-----V-----VL 165
 DB 1501 VELLDRGQVLHLKNARNNDKGRYQCTVSNAAAGQAKDIKLTIIYIPPSIKGNVTTDISVL 1560
 QY 166 -----H----- 168
 DB 1561 INSILKLECKTRGLPMPALTWYKDGQIPMSSSQALYIDKQYLHIIPRAQVSDSATYCHV 1620
 QY 169 -----E-S-----E-----VFI-----C--A-----W----- 177

Db 1621 ANVAGTAEKSFHVDVYPPMIEGNLATPLNKQVVIASHLTLECNAGNPSPILTTLKDG 1680
QY 178 -N- - - - -P-V-SD- 182
Db 1681 PVKANDFRIEAGKKLEIMSAQEIQRQYICVATSVAGEKEIKYEVVDVLVPPAIBEGDE 1740
QY 183 -LL- -VSGS- -G- -D- -S- 191
Db 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKDGQLIDBERDGFILLNGRKLVIQAQVSN 1800
QY 192 -TA- - - - -R- - - - -I-W 196
Db 1801 GLYRCMAANTAGHKKEFEVTVHPPTIKSSGLSERVVVYKPVALQCIANGIPNSITW 1860
QY 197 -NL- -S- - - -E- - - -NS- -T- - - -S 204
Db 1861 LKDDQPNTAQNKKQSSGRVLQIAKTLEDAGRYTCVATNAAGETQQHQLHVHPPS 1920
QY 205 -G- -PTQL- -V- - - -L-R- - - - -212
Db 1921 LEDAGKMLNETVLSNPVQLCKAAGNPVPIITWYKNCLLSGTSMTFLNRQIIDIES 1980
QY 213 -H- - - - -C- - - - -I- 215
Db 1981 AQISDAGIYKCAINSAGATELFYSQVHVAPISGSNNMVAVVVNNPVRLECEARGIPA 2040
QY 216 -R- - - - -E- - - - -G- -Q- -D- - - -221
Db 2041 PSITWLKDGSPVSPFNSGLQVLSGGRIALTSTQISDTGRYTCVAVNAAGEKORDL 2100
QY 222 -VPSN- - - - -KD- - - - -VT- - -S-L- 231
Db 2101 YVPNMGEBQNVSVLSQAVELLQSDAIPPTTLTWLKGHPHLLKPKGLSISENSVLK 2160
QY 232 -D- - - - -W- -N- -S- - - -EG- -TL- 240
Db 2161 IEDAQVQDTGRYTCATNVAGTEKKYNNIWPVNIIGSGDELQTLVIEGNLSILCE 2220
QY 241 AT- - - - -GS- -Y- -D- - - -G- - - -F- - - -A- 249
Db 2221 SSGIPPNLWKKGSPVLTDSMGRVRLSGGRQLQISIAEKSDAALYSCVASNAGTAK 2280
QY 250 -R- - - - -I- - - - -WTKDG- - - - -256
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSISLECEVOGIPPTVTWTKDGHPLIKAGV 2340
QY 257 -NL-AS- - - - -T- - - - -L- - - - -GOH- - - - -265
Db 2341 EILDEGHILQKNIHVSDTGRYVCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K- - - - -G-PI- - - - -FALK- -W- - - - -274
Db 2401 EKNSVLTCEASGIPLPSTTFW- -KOGWPVLSNSVRILSGGRMLRMQTMEDAGQYTC 2458
QY 275 -N- - - - -K-K- - - -GN- - - - -279
Db 2459 VVNAAGEERKIFGLSVLPVPHVIGENTLEDVKVKEQSVTLTCEVTGNPVPEITHKDG 2518
QY 280 -F-I- - - - -L- - - - -S- - - - -AGV- 286
Db 2519 QPLQDEEAHHIISGGRFLQITNVQVPHGTGRYTCCLASSPAGHKRSFSLNVFVSPTIAGVG 2578
QY 287 -D- -K- -T- - - -I- - - - -I-W- - - - -293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPATITWPKDGTPLSNRNIRILPGGRTLQILNAQ 2638
QY 294 -D- - - - -A- - - - -H- - - - -TG- - - - -E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWPGLSPEVKIKVNTLTLECE 2698
QY 300 A- - - - -K- - - - -Q- - - - -Q- - - - -303

Db 2699 AVAIPASLSWYKDGQPLKSDDDHVNIAANGHTLIQKEAQISDTGRYTCVASNIAGEBELD 2758
QY 304 F- - - - -P-F- - - - -H- 307
Db 2759 FDNIOVPPSFQKLWEIGNMLDTRNGEAKDVIIINNPIISLYCETNAAPPTTLTWYKDGHP 2818
QY 308 -SA- - - - -P- - - - -A- - - - -LDV- - - - -D 315
Db 2819 LTTSSDKVLILPGGRVLQIIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878
QY 316 -WQ- - - - - - - - - -SN- - - - -319
Db 2879 LPPEVTVLNKSALIBCLSSGSPAPRNSWQKQGPLLEDDHHKFLSNGRILQILNTQITD 2938
QY 320 -NT- - - - -FAS- -C- - - -S 326
Db 2939 IGRYVCVAENTAGSAXKYFNLNVHVPSPVIGPKSENLTVVNNFISLTCEVSGFPDPDLS 2998
QY 327 -T- - - - -D- - - - -MCI- - - - -331
Db 2999 WLKNXQPIKLTNTLIVPGGRTLIIRAKVSDGGEYTCIAINKAGESKKFSLTVVPPS 3058
QY 332 -H- - - - -V- -C- - - -K- - - -L- -GQ- - -338
Db 3059 IKDHDSESLSVNVREGTSVLECESNAVPPVITWYKNGRMITESTHVEILADGOMLHI 3118
QY 339 -D- - - - -R- - - - -P-IK- - - - -343
Db 3119 KKAESDTCQYVCRAINVAGRDDKNFNLNVPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 -T- - - - -F- - - - -Q- -G- - - -H- - - -T- - - -349
Db 3179 IPPPTIAMLKNYKRIENSLSLEVRILSGSKLQIARSQSDSGNYTCIASNMEGKAKYY 3238
QY 350 -N-B-V-NA- - - - -IKW- -D- -P- - - -T- - - -360
Db 3239 FLISQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTLQIWLKDGKFIASGETERIRV 3298
QY 361 -G- -NL- -A- -S- - - -C- -S- - - -D- - - -D- - - -MTL- - -373
Db 3299 SANGSTLNIYALTSDTGKYTCVATNPAGEEDRIFNLVYVTPTRIGNKDEAKLMTYVD 3358
QY 374 -K- - - - -I-W- - - - -SM- - - - -378
Db 3359 TSINIECRXTGTPPPQINLKNGLPLSLSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418
QY 379 -K- -Q- - - -DN- - - - -CV- - - - -384
Db 3419 RAGVDNKHYNLQVFAFPNNDNSMGTEETVLKGSSTSMACITDGTAPSMAWLRDQPLG 3478
QY 385 -H- - - - -D- -L- - - -QA- - - - -H-N- - - -KE 393
Db 3479 LDAHLTVSTHGMVLQLLKARTEDSGKYTCIASNEAGEVSKHFLKVLPEPHINGSEHEE 3538
QY 394 I- - - - -YT- - - - -I- -K- -W- - - -N- - - - -399
Db 3539 ISVIVNNPLELTIASGIPAPKMTWMDKGRPLPQTDQVQLGGEVLRISTAQVEDTRY 3598
QY 400 -SPTG- - - - -P- - - - -GT- - - - -N- - - - -407
Db 3599 TCLASSPAGDDKEXLVRVHVPPNIAGTDEPRDITVLRNRQVLTLECKSDAVPPPVITWLR 3658
QY 408 N- - - - -P- - - - -N- - - - -ANLM-LAS- -A- -S- -F- - - -420
Db 3659 NGERLOATPRVILSGGRYLQINNADLGTANTYTCVASNIAGKTTREFILTNNVPPNIKG 3718
QY 421 -DSTV- - - - -RL-W- -D- -V- - - -DR- - - - -G- - - -I- - - -433
Db 3719 GPQSLVILLNKSTVLECIAEGVPTPRIITWRKQDAGVLGNHARVSIILENGLHQSAHVTD 3778
QY 434 -C- - - - -I- - - - -H- - - - -TL- - - - -T- - - -439
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHPPSPAPGTNNMTVIVNVQTTLACEATGIPKPSINW 3838

QY 48 A--L--I--S--IIO--KG-- 56
Db 481 SVSLDIKTVLSDGFEYCIASAGTGRAQTFDFSEPPVIVQNNVTVTPGERAVLT 540
QY 57 L--Q--Y--V--E-- 61
Db 541 CLIIASVDNLTQWRNDVRVLAEPARTLANLSLELKSVPKNDAGEYCHWVSSEGGSS 600
QY 62 A--EVS--NE-- 68
Db 601 AASVPLTVQPPKVTVMKQSTGSEVSIMCAGYKPKIAWTVMDFIVGSHRYM 660
QY 69 DGTLP--G--R--P-- 78
Db 661 TSDGTLFKNAAPKADAGIYGLASAGTDKQNSTLRYIEAPKLMVVOSELLVALGDIIV 720
QY 79 E--S--L--S--LI--DA--V-- 87
Db 721 MECKTSGIPPPQVWFKGDLRLPSTFLIIDPLGLLKIQETQDLADGDTVCVAINAAGR 780
QY 88 M--P--DV--VO--T-- 94
Db 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLFCYVQGYPEPTIKWRRLDNMFIKRPSPS 840
QY 95 R-- 95
Db 841 VSSIQLRTGALFILNLWASDKGTIYCEAENQFGKIQSETTVTVTGLVAPLIGISPSVAN 900
QY 96 QO--A--Y--R--D--KL--AQ-- 105
Db 901 VIEGQOULTPCTLLAGNPIERRWIKNSAMLLQNPYITVRSGLSHIERVOLGDGEYTC 960
QY 106 OH--A-- 108
Db 961 VASNAVAGTNNKTSVVVHVLPITQHQOILSTIEGIPVTLPCASGNPKPSVIMSKKGL 1020
QY 109 A--A--A--A--A-- 114
Db 1021 ISTSSAKFSAGADSLVVSPEGESEYVCTATNTAGYAKRVQVTVYVPRVFGDLRG 1080
QY 115 A--A-- 116
Db 1081 LSQDKPVEISVLAGEEVLTPCEVKSUPPIITWAKETQLISPFSPRHTFLPSSGSMKITET 1140
QY 117 ATN--Q--Q--GSA--K--N-- 126
Db 1141 RTSDSGMYLCVATNIAGNVTOAVKLVNHVPPKIQRGPKHLKVQVQORVDIPCAQGTPLP 1200
QY 127 GE--N--T--A--N--G-- 133
Db 1201 VITWSKGSNMLVDGHHVSNPDGTLSDOATPSDAGIYTCVATNIAGTDEITLHVQE 1260
QY 134 E--E--N--GA--H-- 139
Db 1261 PPTVEDLEPPYNTTFQERVANORIEPCPAKGPPTIKWLHNGRELTGREGISILEEG 1320
QY 140 T--IA--N--N--HT--DM-- 148
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKDKQVSNVSVLLNQLTNL 1380
QY 149 M--E-- 150
Db 1381 FCEVEGTPSPFIWYKDNVQVTESSITQTVNNGKILKLFRATPEDAGRYSCKAINIAGTS 1440
QY 151 VD--G--DV--E--I--P-- 158
Db 1441 QKYFNIDVLPPTIIGTFPKEVSVNLNRDVALEQVKGTFPFIHWFKDGKPLFLGDPN 1500
QY 159 SN--K--A--V-- 165
Db 1501 VELLDRGOVLHLKNARRNDKGRYQCTVSNAAQKADIKLTIYIPPSIKGNVTTDISVL 1560

QY 166 RG--H-- 168
Db 1561 INSILKLECKTRGLPMPAITWKDQOPIMSSSOALYIDRKQYLHIIPRAQVSDSATYCHV 1620
QY 169 E--S--E--VF--C--A--W-- 177
Db 1621 ANVAGTAESFHVVVVPPMIENGLATPLNKQVVIASHLTLECNAAGNFSPILTWLKGV 1680
QY 178 N--P--V--SD-- 182
Db 1681 PVKANDNPRIEAGGKLEIWSAQEIDRGQYICVATSVAGEKEIKYEVDLVLPVPAIEGGDE 1740
QY 183 LL--VSGS--G--D--S-- 191
Db 1741 TSFYIVMNNLLELDCHVTGSPPTIMWLKQGLIDERDGFKILLNGRKLVIQAQVSNT 1800
QY 192 TA--R--I--W-- 196
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSSGLSERVVVKYKPVALQCIANGIPNSITW 1860
QY 197 NL--S--E--NS--T--S-- 204
Db 1861 LKDDQPVNTAQGNLKIQSSGRVLIQIAKTLLLEDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920
QY 205 G--PTOL--V--L--R-- 212
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVTVITWYKDNCLLSGSTSMTFLNRQOIIDIES 1980
QY 213 H--C--I-- 215
Db 1981 AQISDAGIYKCVAINSAGATELFYSLQVHVAPISGSNNMVAVVNNVPRLECEARGIPA 2040
QY 216 R--G--E--G--Q--D-- 221
Db 2041 PSLTWLKDGPVSSFNGLQVLSGGRILALTSTQISDTGRYTCVAVNAAGEKORDILRV 2100
QY 222 VPSN--KD--VT--S--L-- 231
Db 2101 YVPPNIMGEQNVSVLISQAVELLQSDAIPPTLTWLDKGHPHLLKPKGLSISENSVLK 2160
QY 232 D--W--N--S--EG--TL-- 240
Db 2161 IEDAQVDTGRYTCBATNVAGTEKKNVNNVWPPNIGGSDELQTLQTVIEGNLISLCE 2220
QY 241 AT--GS--Y--D--G--F--A-- 249
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRIILSGGRLQISIAEKSDAALYSCVASNVAGTAK 2280
QY 250 R--I--WTKDG-- 256
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSISLECEVQGIPTPTVWMDKGHPHLLKAKGV 2340
QY 257 NL--AS--T--L--GQH-- 265
Db 2341 EILDEGHILQKNIHVSIDGRYVCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPEISVV 2400
QY 266 K--G--PI--FALK--W-- 274
Db 2401 EKNSVLTCEASGIPLPSTTWP--KGNVPSLSNSVRIILSGGRMLRLMQTTMEDAGQYTC 2458
QY 275 N--K--K--GN-- 279
Db 2459 VVRNAAAGEBKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVEITWHKDG 2518
QY 280 F--I--L--S--AGV-- 286
Db 2519 QPLQEDAEHHIISGGRFLQITNVQVPHTRGYTCCLASSPAGHSRSPSLNVFVSPPTIAGV 2578
QY 287 D--K--T--I--W-- 293
Db 2579 SGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLESNRNIRILPGRTLQILNAQ 2638
QY 294 D--A--H--TG--E-- 299

Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGGLSPKVEKVIKVNNTILECE 2698
QY 300 A-----K-----Q-----Q----- 303
Db 2699 AVAIPASASLWYKDGQPLKSDHVNIAANGHTLIQKEAQISDGTGRVTCVASNIAGEBELD 2758
QY 304 F-----P-F-----H-----H- 307
Db 2759 FDNVIOVPPSFQKLWEIGNMLDTCRNGEAKDVIIINNPIISLYCETNAAPPPTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTTSSDKVLIILPGRVLOIPRAKVEDAGRYTCVAVNEAGESLQYDVRVLVPIIEGANS 2878
QY 316 -----WO-----SN----- 319
Db 2879 LPBEVTVLVNKSALIECLSSGSPAPRNSWQKDGQPLEDDHKKFLSNGRILQILNTQITD 2938
QY 320 -----NT-----PAS--C-----S 326
Db 2939 IGRYVCVAENTAGSAKYNLNVHVPSPVIGPKSENLTVVVNNFISLTCEVSGFPDPLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKNXQPIKNTNTLIVPGGRTLIQIRAKVSDGGEYTCIAXNAGESKKFSLTVVYPPS 3058
QY 332 --H-----V--C-----K-----L-GO----- 338
Db 3059 IKDHSLSVNVNREGTSVLECESNAVPPVITWYKNGRMITESPHVEILADGQMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KGAESDGTGOVCRAINVAGRDKNFHLNVYVPPSIEGPEREIVETISNPVTLTCDATG 3178
QY 344 --T-----F-----Q--G-----H-----T----- 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGKLQIARSOHSDSGNYTCIASNMEGKAQYY 3238
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360
Db 3239 FLSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIQLKDGKPIASGETERIRV 3298
QY 361 --G--NLL-A-S-----C-S-----D-----D-----MTL-- 373
Db 3299 SANGSTLNIYGALTSDTGKTCVATNPAGEEDRIFNLNVYVTPTRGNKDEAEKLMTYD 3358
QY 374 --K-----I-W-----SM----- 378
Db 3359 TSINIECRXTGTPPPQINLWLNGLPLPLSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418
QY 379 --K--Q-----DN-----CV----- 384
Db 3419 RAGVDNKHYNLQVAPPNMDSNGTEBITVLKGSSTSMACITDGTAPASMAWLRDGOPLG 3478
QY 385 --H-----D--L-----QA-----H-N-----KE 393
Db 3479 LDAHLTVSTHGMVLQLKAETEDSGKYTCIASNEAGEVSKHFLKVLPEPPHINGSSEHEE 3538
QY 394 I-----YT-----I--K--W----- 399
Db 3539 ISVIVNPNBELTCIASGIPAKMTWMDGRPLPQTDQVQTLGGGEVLRISTAQVEDTGRY 3598
QY 400 -----SPTG-----P--GT-----N----- 407
Db 3599 TCLASSPAGDDKEYLVRVHPNPIAGTDEPRDITVLNRQVLTLECKSDAVPPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420
Db 3659 NGERLOATPRVILSGRYLOINNADLGDYANTCVASNIAGTKTTFEFLTVNVPNNIKG 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433

Db 3719 GPQSLVILLNKSTVLECIAEGVPTPRITWRKDGAVLGNHARYSILENGFLHIQSAHVTD 3778
QY 434 -----C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHPPSIAPGFTNMVTIVNVQTTLACEATGIPKPSINW 3838
QY 440 -K--H-----O-E-----P-V-----Y----- 446
Db 3839 RKNGHLLNVDOQNSYRLLSSGSLVIISPSVDDTATYEYTVNGAGDDKRTVDLTVOVPP 3898
QY 447 SVA-----F-S-----P-D----- 453
Db 3899 SIADPTDFLVTKHAPAVITCTASGVPPFSIHWTKNGIRLLPRGDGYRILSSGAIBLAT 3958
QY 454 -----GRY-----L-----S--S-- 461
Db 3959 QLNHAGRYTCVARNAGSAHRHVTLVHPPVPIQPOSELHVLINPNPILLPCEATGTPSP 4018
QY 462 F-----D-----KCV----- 466
Db 4019 FITWQKEGINVNTSGRNHVLPSGGLQISRAVREDAGTVMCAQNPAGTALGKILNVQV 4078
QY 467 -----H-----I-W----- 469
Db 4079 PPVISPHLKEYVIAVDKPTILSCADGLPPPDIHTWKDGRAIVESIRQVRLSSGSLQIAF 4138
QY 470 -----N-----T-----Q----- 472
Db 4139 VOPGDAGHYTCMAANVAGSSSTSKLTVHVPPIRSTEGHYTVNENSOAILPCVADGPT 4198
QY 473 -----T-----G-----A-----L-VH 478
Db 4199 PAINWKKONVLLANLLGKYTABPYGELILENVVLEDSGFYTCVANNAGEDHTVSLTVH 4258
QY 479 -----S-----Y-----R 481
Db 4259 VLPTTFELPDVSLNKGEQLRLSCKATGIPKLWTNNNIIPAHFDVSNVGHSELVIER 4318
QY 482 -----GT-----G--G-----IF-----E--V--C----- 490
Db 4319 VSKEDSGTYVCTAENSVGFKATGTFYVYKPPVFKGDPFNSNWIPLGNAILNCEVKGDP 4378
QY 491 -----WN-----A-----AGD----- 496
Db 4379 TPTIQNRKGVDIIEISHRIROLNGSLAIYGTWNEDAGTYTCVATNEAGVVERSMSTLQ 4438
QY 497 -----K-V--G-----A-----SAS----- 503
Db 4439 SPPIITLPEVTVINAGGKIILNCQATGBPQPTITWSRQGHISISWDDRVNVLNNSLYIA 4498
QY 504 -----D-----GSV-----C-V----- 509
Db 4499 DACKEDTSBECVARNLMSVLVRVPIVQVHGGFQSWAWRACSVTCGKGIQKRSLCN 4558
QY 510 --L-----DL--R-----K 514
Db 4559 QPLPANGKPKCGSDLEMRNCQNK 4582

RESULT 7

US-10-023-634-93

; Sequence 93, Application US/10023634

; Publication No. US20030236389A1

; GENERAL INFORMATION:

; APPLICANT: Shimkets, Richard A

; APPLICANT: Colman, Steven D

; APPLICANT: Spyttek, Kimberly A

; APPLICANT: Ballinger, Robert A

; APPLICANT: Guo, Xiaojia

; APPLICANT: Tchernev, Velizar T

; APPLICANT: Shenoy, Suresh G

; APPLICANT: Li, Li

; APPLICANT: Ellerman, Karen

APPLICANT: Zerhusen, Bryan D
APPLICANT: Patturajan, Meera
APPLICANT: Casman, Stacie J
APPLICANT: Boldog, Ferenc
APPLICANT: Gusev, Vladimir Y
APPLICANT: Burgees, Catherine E
APPLICANT: Edinger, Shlomit R
APPLICANT: Gangolli, Esha A
APPLICANT: Malyankar, Uriel M
APPLICANT: Gunther, Erik
APPLICANT: Smithson, Glennda
APPLICANT: Millet, Isabelle
APPLICANT: Gerlach, Valerie
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
FILE REFERENCE: 21402-221
CURRENT APPLICATION NUMBER: US/10/023,634
CURRENT FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: 60/256,025
PRIOR FILING DATE: 2000-12-15
PRIOR APPLICATION NUMBER: 60/265,163
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/272,929
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/274,864
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/276,688
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/277,880
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 60/286,409
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/309,246
PRIOR FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/315,600
PRIOR FILING DATE: 2001-08-29
NUMBER OF SEQ ID NOS: 132
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 93
LENGTH: 5636
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (3003)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
OTHER INFORMATION: specification.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (3041)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
OTHER INFORMATION: specification.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (3367)
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
OTHER INFORMATION: specification.
US-10-023-634-93

Query Match 71.6%; Score 2627.8; DB 14; Length 5636;
Best Local Similarity 9.6%; pred. No. 9.8e-30;
Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;

QY 1 M-S-----I-----SS---D-----EV----- 8
DB 1 MISWEVHTVFLPALLYSLAQPASQSEIRAEFFEGASTLAFVDFVTGSMYDVLQVI 60
QY 9 -----NF-LV----- 12
DB 61 EGASKILETSLKRPKPLNFALPHDPEIGPVITITDPKFKQYELRELYVGGGDCPE 120
QY 13 -----Y-----R---Y-----LQ----- 17

DB 121 MSIGAIIKIALEISLPGSFYIVFTDARSKDYRLTHEVLQIIQQKQSVQVFLTGDCDDRTH 180
QY 18 -----E-----SG-F-----S-H-S-----A-----F 26
DB 181 IGYKVEEIIASTSSGOVFLDKKQVNEVLKQVVEEAQKSVHLLSTDHLEQAVNTWRIFP 240
QY 27 -----T-----FG-----I-----E----- 31
DB 241 DPSLKEVTVLSGSPSPMIEIRNPLGKLIKKGFGHELLNTHNSAKVNVNVEPEAGMWTVK 300
QY 32 ---S-H-----I-----S-Q----- 36
DB 301 TSSGRHSVRITGLSTIDFRAGFSRKPTLDFKTKTSRVPQGIPTVYLLNTSGISTPARID 360
QY 37 -----S-----NI-----N-----G----- 41
DB 361 LLELLISGSSSLKTPVKYYPHRKPYGIWNISDFVPNEAFFLKVTKGYDKDDYLFQRVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
DB 421 VSFSSIVDPAPKVTMEKTPGYVLQPGQIPCSVDLSLPFTLSFVRNGVTGLGVDOYLKESA 480
QY 48 -----A-----L-----I-----S-----IIQ-----KG----- 56
DB 481 SVSLDIKAVTSLDSDEGFYECIAVSAGTGRAQTFFDVSEPPPIQVPPNNVTVTGERAVLT 540
QY 57 -----L-Q-----Y-----V-E----- 61
DB 541 CLIIISAVDYNLTWQRNDVRVLAEPARITLANLSLELSKVKFNDAGEVHCMSSEGSS 600
QY 62 -A-----EVS-----NE----- 68
DB 601 AASVFLTVQEPKVTVMKQNSFTGGSEVSIMCSATGYPKPKIAVTNDMFIVGSHRYRM 660
QY 69 ---DGLF-----D-----G-----R-----P-----I--- 78
DB 661 TSDGTLFIKNAAPKADAGIYGCILASNSAGTDKQNSTLRYIEAPKLMVQSELLVALGDITV 720
QY 79 -E---S-----L-S-LJ-----DA-----V----- 87
DB 721 MECKTSGIPPPQVQKWFKGDLERPSTFLIIDPLGLLKIQETQDLDAGDYTCVAINAGR 780
QY 88 -----M-----P-DV-----VQ-----T----- 94
DB 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRLLDNMFISRPFS 840
QY 95 -----R----- 95
DB 841 VSSISQRLTGALFILNLWASDKGTIYCEAENQFGKIQSETTIVTGLVAPLIGISPSVAN 900
QY 96 ---QO-----A-----Y---R-D-KL---AQ----- 105
DB 901 VIEGQQLTLPCTLLAGNPPIERRWIKNSAMLLQNPYITVRSQGLSHIERVOLQDGGEYTC 960
QY 106 -----QH-----A----- 108
DB 961 VASNVAGTNKTTVVVVHPTIQHQQILSTIEGIPVTLPCCKASGNPKPSVIWKKGEL 1020
QY 109 -----A---A-A-----A---A-A----- 114
DB 1021 ISTSSAKPSAGADGSLYVVSPEGESGEYVCTATNTAGAKRKVQLTVYVPRVFGDLRG 1080
QY 115 -----A-----A----- 116
DB 1081 LSQDKPEIVSLAGEEVTLPCEVKSUPPIITWAKETQLISPPSPRHTFLPSSMKITET 1140
QY 117 -----ATN-----Q-----Q-GSA-K-----N----- 126
DB 1141 RTSDSGMYLCVATNIAGNVTAQVKNVHPVKIQRGPKHLKVQVGQGVQVDIPCNAAQGTPLP 1200
QY 127 -----GE---N-----T---A-N-G----- 133
DB 1201 VITWSKGGSTMLVDGSHHVSNDGTLSIDQATPSDAGIYTCVATNIAGTDEBITLHVQE 1260

134 QY -----E-----E-N-----GA-----H-----139
1261 Db PPTVEDLEPPYNTTFQERVANORIEFPCKAPKPTIKWLHNGRELTGREPGISILERG 1320
140 QY T-----IA-----N-----N-----HT-----DM-----148
1321 Db TLLVIASVTPYDNGEYICVAVNAGTTERKYNLKVHVPVVIKQBSNVSVLLNQLTNL 1380
149 QY -----M-----N-----E-----150
1381 Db FCEVEGTPSPINIMYKDNVQVTESSITQVNNKGILKLPRATEDAGRYSCKAINIAGTS 1440
151 QY -----VD-----G-----DV-----E-----I-----P-----158
1441 Db QKYFNIDVLVPPTIIGTFNFKPKEVSVVLRDVALEQCVKGTGTPFDIHWFKDGKFLGLDPN 1500
159 QY -----SN-----K-----A-----V-----VL-----165
1501 Db VELLDRGQVLHLKARNRNDKRYOCTVSNAGKQAKDIKLTIVIPPSIKGNVTTDISVL 1560
166 QY -----RG-----E-----H-----168
1561 Db INSILKLECKTRGLPMPAITWYKDGQPIMSSQALVIDKQYLHI PRAQVSDSATYTCV 1620
169 QY -----E-----S-----E-----VPI-----C-----A-----W-----177
1621 Db ANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECNAAGNPSPILTLKDG 1680
178 QY -----N-----P-----V-----SD-----182
1681 Db PVKANDNFRIEAGGKLEIMSAQEI DRGQIVCVATSVAGEKEIYEVVDVLVPAIEGDE 1740
183 QY -----LL-----VSGS-----G-----D-----S-----191
1741 Db TSVFIVMNNLLELDCHVTGSPPTIMWLKQGLIDERDGFILLNGRKLVIQAQVSNT 1800
192 QY -----TA-----R-----I-----W-----196
1801 Db GLYRCMAANTAGDHKHEFEVTVHVPPTIKSGLSERVVVVKYKVALQCIANGIPNPSITW 1860
197 QY -----NL-----S-----E-----NS-----T-----S-----204
1861 Db LKDDQPVNTAGNLKIOSGSRVLOIAKTLLDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920
205 QY -----G-----PTOL-----V-----L-----R-----212
1921 Db LEDAGKMLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSGSTMFLNRGOIIDIES 1980
213 QY -----H-----C-----I-----215
1981 Db AQISDAGYKCVAINSAGATELFYSLOVHVAPSIGSGNNMVAVVNNPVRLCEARGIPA 2040
216 QY -----R-----E-----G-----Q-----D-----221
2041 Db PSLTWLKDGPVSSFSNGLQVLSGRILALTSTQIDTGRYTCVAVNAAGEKQORDILRV 2100
222 QY -----VPSN-----KD-----VT-----S-----L-----231
2101 Db YVPEINMGEQNSVLISQAVELLQSDAIPPTTLTWLKDGHPLLRKPLKLSISENSRVLK 2160
232 QY -----P-----W-----N-----S-----EG-----TL-----240
2161 Db IEDAQVQDTGRTCEATNVAGTEKKNVNNIWPVPPNIGGSDDELTLQTLVIEGNLISLCE 2220
241 QY AT-----GS-----Y-----D-----G-----F-----A-----249
2221 Db SSGIPPPNLIWKKGSPVLTDMSGRVRLSGRQLQISIAEKSDAALYSVCVASNAGTAK 2280
250 QY -----R-----I-----WTKG-----256
2281 Db KEYNLQVYRPTITNSGSHPTIIVTRCKSISLECEVOGIPPTVTWVKDGHPLIKAGV 2340

257 QY -----NL-----AS-----T-----L-----GQH-----265
2341 Db EILDEGHILQKNHIVSDTGRYVCVAVNAGMTDKKYDLSVHAPPSIIGNHSPENISVV 2400
266 QY -----K-----G-----PI-----FALK-----W-----274
2401 Db EKNSVLTCEASGIPLPSTWF-----KDGMPVLSLSNVRLSGRMLRLMOTTMEDAGQYTC 2458
275 QY -----N-----K-----K-----GN-----279
2459 Db VVRNAGEERKIFGLSVLPVPHIVGENTLEDVKVEKOSVTLTCEVTGNPVPPIETHKDG 2518
280 QY -----F-----I-----L-----S-----AGV-----286
2519 Db QPLQDEAHIIISGGFLOITNVQVPHTCRYTCLASSPAGHKSRFSLSNVFVSPTIAGVG 2578
287 QY -----D-----K-----T-----I-----I-----W-----293
2579 Db SDGNPEDVTILNSPTSLVCEAYSYPATITFKDGTPLSNRNRILPFGRTLQILNAQ 2638
294 QY -----A-----H-----TG-----E-----299
2639 Db EDNAGRYSCVATNEAGEMIKHYEVKYIPPIINKGDLWPGGLSPKEVKIKVNTLTLECE 2698
300 QY -----A-----K-----Q-----Q-----303
2699 Db AYAI PASLSWYKDGQPLKDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758
304 QY -----P-----P-----H-----307
2759 Db FVNIQVPPSFOKLWBEIGNMLDTRNGEAKDVIIINNPISLYCETNAAPPPTLTWYKDGHP 2818
308 QY -----SA-----P-----A-----LDV-----D-----315
2819 Db LTSSDKVILIPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPILIEGANS 2878
316 QY -----WQ-----SN-----319
2879 Db LPEEVTVLNKSALIECLSSGSPAPRNSWQKQPLLEDHKKFLSNGRILQILNTQITD 2938
320 QY -----NT-----FAS-----C-----S-----326
2939 Db IGRYVCVAENTAGSAKXYFNLNVHVPVSPVIGPKSENLTVVVNNFISLTCEVSGFPDLS 2998
327 QY -----T-----D-----MCI-----331
2999 Db WLKXQOPIKLTNTLIVPGRTLQIIIRAKVSDGGEYTCIAINXAGESKKFSLTVVYVPS 3058
332 QY -----H-----V-----C-----K-----L-----GO-----338
3059 Db IKDHSLSLVNVNREGTSVLSCESNVAVPPVITWYKNGRMITESTHVEILADGQMLHI 3118
339 QY -----D-----R-----P-----IK-----343
3119 Db KKAESDVTGOYVCRAINVAGRDDKNFHLNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178
344 QY -----T-----F-----Q-----G-----H-----T-----349
3179 Db IPPPTIATLWLNKYRIENSLSLEVRILSGSKLOIARSQHSDSGNVYTCIASNMEGKAQKY 3238
350 QY -----N-----E-----V-----NA-----IKW-----D-----P-----T-----360
3239 Db FLISIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIOWLKDGKPIASGETERIV 3298
361 QY -----G-----NLL-----A-----S-----C-----S-----D-----D-----MTL-----373
3299 Db SANGSTLNIYALTSDTGKTCVATNPAGEEDRIFNLNVVVTPIRGNKDEAEKLMTYVD 3358
374 QY -----K-----I-----W-----SM-----378
3359 Db TSINIECRXTGTPPPQINWLNKGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418
379 QY -----K-----Q-----DN-----CV-----384

481 SVSLDIKVTLSDEGVEICLAVSSACTGRAQTFDDVSEPPVIVQVNNVTVTPGERAVLT 540
QY - - - - - L - Q - - - - - Y - - - - - E - - - - - 61
541 CLIIISAVDYNLTWORNDVRVLAEPARIRTLANLSLELKSVEKFNDAEYCHMVSSSEGGSS 600
QY - - - - - A - - - - - EVSI - - - - - NE - - - - - 68
601 AASVFLTVQEPKVTWMPKQNSGTGSEVSIMCSATGYPKPKIANTVNDMFIVGSHRYM 660
QY - - - - - D - - - - - G - - - - - R - - - - - P - - - - - I - - - - - 78
661 TSDGTLFIKNAAPKDAIGYCLASNAGTDKQNSTLYRIEAPKLVVQSELLVALGDITV 720
QY - - - - - S - - - - - L - S - LI - - - - - DA - - - - - V - - - - - 87
721 MECKTSGIPPPQVKFKGDLERFSTFLIIDPLLLKIKIQTQDLDDAGDYTCVAINAAGR 780
QY - - - - - M - - - - - P - DV - - - - - VQ - - - - - T - - - - - 94
781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMFIKRPFS 840
QY - - - - - R - 95
841 VSSISQIRTGALFILNWDKGTICYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVAN 900
QY - - - - - Q - - - - - A - - - - - Y - - - - - R - D - KL - - - - - AQ - - - - - 105
901 VIBEQQLTLPCTLLAGNPIPERRWKNSAMLLQNPYITVRSDSLHIERVQLODGEYTC 960
QY - - - - - OH - 108
961 VASVAGTNNKTSVVVHVLPITQHQOILSTLEGIPVTLPCASGNPKPSVWSKKGEL 1020
QY - - - - - A - - - - - A - - - - - A - - - - - A - - - - - - - - - - - - - - - 114
1021 ISTSSAKFSAGAGSLVVSPEGESYVCTAINTAGYAKRVQVLTVYVRPVFGDLRG 1080
QY - - - - - A - - - - - A - - - - - A - - - - - A - - - - - - - - - - - - - - - 116
1081 LSQDKPVEISVLAGEEVLTPCEVKSLLPPIITWAKETQLISPPSPRHTFLPSGSMKITET 1140
QY - - - - - ATN - - - - - Q - - - - - Q - GSA - K - - - - - N - - - - - - - - - - 126
1141 RTSDSGYLCVATNIAGNVQAVKLVNHPVKIQRGPKHLKVQGVQVDPICNAOQTPLP 1200
QY - - - - - GE - - - - - N - - - - - T - - - - - A - N - G - - - - - - - - - - 133
1201 VITWSKGSMTLVDEGHHVSNPDGTLISIDQATPSDAGIYTCVATNIAGTDEBITLHVQE 1260
QY - - - - - E - - - - - E - - - - - N - - - - - GA - - - - - H - - - - - - - - - - 139
1261 PPTVEDLEPPYNTTFOERVANORIEPPCPAKGTPTIKWLHNGRELGTGREPGISILEEG 1320
QY - - - - - IA - - - - - N - - - - - N - - - - - HT - - - - - DM - - - - - - - - - - 148
1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKKEQVSNVSVLLNQLTNL 1380
QY - - - - - M - 150
1381 FCEVEGTPSLIMYKONVQVTSSTIQTVNNKILKLFRATPEDAGRYSCKAINIAGTS 1440
QY - - - - - VD - - - - - G - - - - - DV - - - - - E - - - - - I - - - - - - - - - - P - 158
1441 QKYNIDVLVPTTIIGNFPKSVNLNRDVALEQVKGTFPPFDIHFWDKGLFLGDPN 1500
QY - - - - - SN - - - - - K - A - - - - - V - - - - - V - - - - - - - - - - - - - - - 165
1501 VELLDRQVHLKVARNDKGRYQCTVSNAGKQAKDKLTIYIPPSIKGNVTTDISVL 1560
QY - - - - - RG - 166
1561 INSLIKLECKTRGLPMPAITWYKDGQPISSSQALYIDKQYLHPIRAQVSDSATYVTCVH 1620
QY - - - - - E - S - - - - - E - - - - - VFI - - - - - C - A - - - - - W - - - - - 177
1621 ANVAGTAKSFHVDVYVPPMIEGNLATPLNKQVIAHSLTLECNAGNPSPLTLWLKQGV 1680
QY - - - - - N - 182
1681 PVKANDNFRIEAGGKKLEIMSAQEIDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740
QY - - - - - LL - - - - - VSGS - - - - - G - - - - - D - - - - - S - - - - - 191
1741 TSYFIWMNNLLELDCHVTGSPPTIMMLKQQLIDERDGFKILLNGRKLKLVIAQVSVNT 1800
QY - - - - - TA - - - - - - - - - - - R - 196
1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSSGLSERVVVYKPVALQCIANGIPNPSITW 1860
QY - - - - - NL - - - - - S - - - - - E - - - - - NS - - - - - T - - - - - S - - - - - 204
1861 LKDDQPVNTAQNLIKIQSSGRVLIQIAKTLLEDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920
QY - - - - - G - - - - - PTOL - - - - - V - 212
1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLSGSTMTFLNRQIIDIES 1980
QY - 215
1981 AQISDAGIYKCVAINSAGATELFYSLQVTVAPISGSNNMVAVVNNPVRLCEARGIPA 2040
QY - - - - - R - - - - - - - - - - - E - G - - - - - G - Q - - - - - D - - - - - 221
2041 PSITWLKDGSPVSSFNGQLVLSGGRILALTSTQISDTGRYTCVAVNAAGEKQORDILRV 2100
QY - - - - - VPSN - 231
2101 YVPPNIMGBEQNVSVLISOAVELLQSDAIPPTTLWLKDGHPLLKKPGLSISENRSVLK 2160
QY - - - - - D - - - - - W - - - - - N - - - - - S - - - - - EG - - - - - TLL - - - - - 240
2161 IEQAQVDTGRYTCBATNVAGTEKKNVNNVNPENIGSGDELTQTLTVEGNLISLCE 2220
QY - - - - - AT - - - - - GS - - - - - Y - D - - - - - G - - - - - F - - - - - - - - - - A - 249
2221 SSGIPPPNLIWKKGSPVLTDSMGRVRIILSGRQLOISIAEKSDAALYSCVASNVAGTAK 2280
QY - 256
2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSI SLECEVQGI PPPPTVWMDKGHPLIKAKGV 2340
QY - - - - - NL - AS - - - - - T - - - - - L - 265
2341 EILDEGHILQKNHIVSDTGRYVCVAVNAGMTDKKYDLSVHAPPSIIGNHRSPIENISVV 2400
QY - - - - - K - - - - - G - PI - - - - - FALK - W - 274
2401 EKNSVLTCEASGIPLPSTTF - - - - - KDGWPVSLNSVRILSGGRMLRMQTTMEDAQOYTC 2458
QY - 279
2459 VVRNAGEERKIFGLSVLPPPHIVGENTLEDVKVKEKQSVTLTCEVTGNPVPBITWHKDG 2518
QY - - - - - F - - - - - I - 286
2519 QPLQEDENHIIISGGRFLQITNVQVPHTRGYTCCLASSPAGHKRSFSLNVFVSPPTIAGVG 2578
QY - - - - - D - - - - - K - - - - - T - - - - - I - - - - - I - W - - - - - - - - - - - - - - 293
2579 SDGNPEDVTILNSPTSLVCEAYSPPTATITWFKDGTPLSNRNRILPGRTOILNAQ 2638
QY - - - - - D - - - - - A - - - - - H - 299
2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGLSPKVEVKIKVNNTLTLECE 2698

QY 300 A-----K-----Q-----Q----- 303
Db 2699 AYAIASLSWYKDGOPKPLSDHVNIAANGHTLQIKEAQSITGRYTCVASNIAGEDEL 2758
QY 304 F-----P-F-----H----- 307
Db 2759 FVNIQVPPSFQKWEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPPTLTWYKDGHP 2818
QY 308 --SA-----P-----A-----LDV-----D 315
Db 2819 LTSSDKVILPGGRVLQIPRAKVEDAGRYTCVAVNEAGDSLOYDVRVLVPPPIEGANS 2878
QY 316 -----WQ-----SN----- 319
Db 2879 LPEEVLVWKSALTECLSSGSPAPRNSWQKQDQPLEDDHFKFLSNGRILQILANTQID 2938
QY 320 -----NT-----FAS--C-----S 326
Db 2939 IGRYVCVAENTAGSAKYFNLVHVPVPSVIGPKSENLTVVVNNFISLTCEVSGFPPDLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKNQOIKLNTLIVPGERTLQIIRAKVSDGEYTCIAINXAGESKKFSLTVVPPS 3058
QY 332 --H-----V--C-----K-----L--GO----- 338
Db 3059 IKDHSESLSVNVREGTSVLECESNAVPPVITWYKNGRMITBSTHVEILLADQMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KKAESDVTQYVCRAINVAGRDDKNFNLVNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 --T-----F-----Q--G-----H-----T----- 349
Db 3179 IPPPTIAMLKNYKRIENSLSLEVRILSGSKLQIARSHSDSNYTCIASNMEGKAQKY 3238
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360
Db 3239 FLISQVPPSVAGAEIPSDVSVLLGENVELVCHANGIPTPLIQWLKDGKPIASGETERIRV 3298
QY 361 --G--NULL-A--S-----C--S-----D-----D-----MTL-- 373
Db 3299 SANGSTLNIYALTSYGTGYTCVATNPAGEEDRIFNLNVVYPTIRGNKDEAEKLMYVD 3358
QY 374 --K-----I-W-----SM----- 378
Db 3359 TSINIECRXTGTPPPQINLWKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418
QY 379 --K--Q-----DN-----CV----- 384
Db 3419 RAGVDNKHYNLQVAPPNNMDNSMGTEETVLKGSSTSMACITDGTAPPSMAMLRDQOPLG 3478
QY 385 --H-----D--L-----QA-----H-N-----KE 393
Db 3479 LDAHLLTVSTHGMVQLLKAETEDSGKYTCIASNEAGEVSKHFLKYLEPPHINGSEHEE 3538
QY 394 I-----YT-----I--K--W----- 399
Db 3539 ISVINNPULETCIASGIPAPKMTWMMKGRPLPQTDQVOTLGGGEVLRISTAQVEDTGRY 3598
QY 400 -----SPTG-----P--GT-----N----- 407
Db 3599 TCLASSPAGDDDEKYLVRVHPVNPNTAGTDEPRDITVLRNROVTLECKSDAVPPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS--A--S--P----- 420
Db 3659 NGERLOATPRVILSGRYLQINNADLGDANTVTCVASNIAGKTTREFILTVNVPNIKG 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433
Db 3719 GPQSLVILLNKSTVLECIAGVPTPRITWRKDGAVLAGNHARYSILENGFLHQSAHVTD 3778

QY 434 --C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHPPPSIAPGPTNMTVIWVQTTLLACEATGIPKPSINW 3838
QY 440 --K--H-----Q-E-----P-V-----Y----- 446
Db 3839 RKGHLLNVNDQNSYRLLSSGSLVSIISFSDDTATYECTVTNGAGDKRTVDLTVQVPP 3898
QY 447 SVA-----F-S-----P-D----- 453
Db 3899 SIADPTDELVTKHAPAVITCTASGVPPFSIHWTKNGIRLLPRGDRYLSSGAIBLAT 3958
QY 454 --GRY-----L--ASG--S-- 461
Db 3959 QLNHAGRYTCVARNAAGSAHRHVLHLVHPPVIQOPSELHVLNPNILLPCEATCTPSP 4018
QY 462 F-----D--KCV----- 466
Db 4019 FITWQKEGINVNTSGRNHVLPSGGLQISRAVREDAGTYMCVAQNPAGTALGKIKLVQV 4078
QY 467 --H-----I-W----- 469
Db 4079 PPVISHPLKEYVIAVDKPTILSCREADGLPPPDIWTKDGRAIVESIRQVRVSSGSLQIAF 4138
QY 470 -----N-----T-----Q----- 472
Db 4139 VQPDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTVNENSQAILPCVADGPT 4198
QY 473 --T-----G-----A-----L-VH 478
Db 4199 PAINWKNVLLANLLGKYTABPYGELILENVLEDGTFYTCVANNAAGEDTHTVSLTVH 4258
QY 479 --S-----Y-----R 481
Db 4259 VLPPTTELPGDVSINKGBQLRSLCKATGIPKLTWTFNNIIPAHFDSVNGHSELVIER 4318
QY 482 --GT-----G--G-----IF-----E--V--C----- 490
Db 4319 VSKEDSGTYVCTAENSUGFVKATGTFVYVKEPPVFGDPSNMIEPLGNAILCEVKGDP 4378
QY 491 --WN-----A-----AGD----- 496
Db 4379 TPTIOMNRKGVDEIESHRIQLNGSLAIYGTVNEDAGDYTCVATNEAGVWVERSMTLQ 4438
QY 497 --K-V--G-----A-----SAS----- 503
Db 4439 SPPIITLBPVETVINAGKIIILNCOATGEPQPTITWSROGHSISWDDRVNVLNNSLYIA 4498
QY 504 --D-----GSV-----C-V----- 509
Db 4499 DAQKEDTSEFECVARNLMSVLVRVPVIVQVHGFSGWSAMRACSVTCGKIQKRSRLCN 4558
QY 510 --L-----DL--R-----K 514
Db 4559 QPLPANGGKPCQSGDLEMENCONK 4582

RESULT 9

US-10-464-368-69
; Sequence 69, Application US/10464368
; Publication No. US20040023356A1
; GENERAL INFORMATION:
; APPLICANT: Krumlauf, Robb
; APPLICANT: Ellies, Debra
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
; FILE REFERENCE: 40716-IP-017
; CURRENT APPLICATION NUMBER: US/10/464,368
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/388,970
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 69

Query Match
Best Local Similarity 9.58; Pred. No. 6.5e-30;
Matches 436; Conservative 55; Mismatches 17; Indels 4081; Gaps 381;

1 MS-----I-----S-----4
1 MSQALLAILTSLGLENVLIIGANQDQHLCDPGEFLCHDHVTCVSQSWLGDGDPDCPD 60
5 -SD-----EV-----N-----F-----10
61 QSDLESITDCEVEIKPLNHIACHGSACVHLKLCNGVVDPCDPDGEHGHQELLPS 120
11 -L-----V-----Y-----R-----Y-L-Q-----17
121 QQLNCQFKAMVFNATRCYCEDGFEVADGRSCKDQDECSIYICISQTCNKTYGVACSC 180
18 -ESG-----F-----S-----H-----23
181 VE-GYMQSDNRCKVKHEPTKAPMLLISSLETIELFYINGSKMTLSSANRNEIHTLD 239
24 -----S-----AF-T-----27
240 FIYSEMICWIESRESSNQLKCOQITKAGRLTDORINSLSQSPQNVQWAFDLTRNIYP 299
28 -----F-G-----IES-H-----Q-----33
300 VDHVSDFRIFVCFNPGSVCTVLISELNPKAIAADPIAGKLPFTDYGVPKIERCDDGM 359
34 -----S-----S-----Q-----36
360 NRTRIVSKAEQPSALALDLNRLVYVDLYDYGVDYQGNRHTIVQGRQVHLYGI 419
37 -----S-NI-----NG-----A 42
420 TVFEDYLYATSDNFNIIRINRNGTDIHSIIKMSARGIRTYQKRTQPTVRSHACEVDA 479
43 -----42
480 YGMPGGCHICLSSSKTKTCRTGPNMGSDGRCKRPKNELFLFYGKRGPIVRGMD 539
43 L-----VP-----P-----A-A-LI-----S-II-----OK-G-----L- 57
540 LNTKIADECMIPENLVNPRALDFAEANYIYFADTTSLIGRQKIDGTERETILKDDLD 599
58 -Q-----Y-----V-----E-A-----E-----VS-IN-----67
600 NVEGIADVITGNLYWTDNGHRKTINVARLEKASQSRKTLLEGGMSPRAIVDPVNGWM 659
68 -----ED-----G-TL-F-----D-----74
660 YWTDWKEDKIDDSVGRLEKAWDGVNRQVFTSKMLWPNGLTLDFHTSLTYWCDAYYDHI 719
75 -----G-R-----P-----I-----E-SL-----81
720 EKVFNGTHRKVVYSGKELNHPFLSHGHVYFWTDYMGSIYFQDLMTNEVTLRHERA 779
82 -----SL-I-----D-----A-V-----87
780 PLFGLQIDPRKQGGNMCRINNGCGTCLAIAPAGRVACADNQLLDENGTTCTFNPEE 839
88 -----MP-----D-V-----V 92
840 IRFHCKPGEFRCKNKHCIQARWKCDGDDCLDGSDEDSVTCFNHSCPDQFQKQNNRCI 899
93 -----QT-----R-Q-Q-----AY-R-----DK- 102
900 PKRWLCDGANDCGSNEDESNQTCRTAQDQFS CGNGRCIPTAWLDCREDDCGQDTDEV 959

QY 103 -----IAQ-----Q-----106
Db 960 ASCEPPTCEPLTQFTICKSGRCISNKHCHDTHDDCCDRSDEVGVHSCLDQDQRCSSGRCI 1019
QY 107 --H-A-----AAAA--A-----114
Db 1020 PGHWACDGDNDGDFSDETHINCTKEARSAGCIGNBQCRPDGNCIPDLWRCDGEKDC 1079
QY 115 -----A-A-AT-----N-----Q-----120
Db 1080 EDGSDEKCGNGTIRLCDHKTKFSRCRGINNAWCDGDVDCDQSDDEEDCDLFCGPP 1139
QY 121 -----Q-----GSAK-----NG-----127
Db 1140 KYPCANDTSVCLQPEKLCNGRKDCPDGSDGDLCDCECSLNGCSNHCSSVVPGRGIVCSC 1199
QY 128 -E-----N-T-----A-----N-GE-----E 135
Db 1200 PEGHQLKDNRTCEIVDYCASHLRCSQVCEQQHMKVCSCYEGWALGTGDSCTSVDSFE 1259
QY 136 -----TD-----M-----M-----E-----150
Db 1260 AFIIISIRHEIRIDLHKGDYSLVPLGRNTIALDFHFNQSLLYWTDVVEDRIYRGLSE 1319
QY 137 --G-A-----H-----T-----IA-N-----N-----H 145
Db 1320 SGVSAIEVVVEHGLATPEGLTVDWIAIGNIYWDNSLDQIEVSKLDGSLRATLIAGAMEH 1379
QY 146 -----DG-D-VEI-----P-----S-----N-KA-----162
Db 1380 PRAIALDPYGLTFTWDMDFANFPRIESASMSGAGRKTIYKDMKTGAMPNGLTVDHFERRI 1439
QY 151 V-----L-----L-----RG-----H-----168
Db 1440 VMTDARSDAIYSAFYDGTNMIIEIRGHEYLSPFAVSLYGVYTDWTDNTLAKANKWT 1499
QY 163 -----VV-----L-----RG-----H-----168
Db 1500 QGNVSVIOKTSQAQPDQIYHPSRQAPNPCAANEGRGPCSHLCLINHRSAACAPHL 1559
QY 169 -----E-----SE-----V-----172
Db 1560 MKLSSDKTKCYEMKFKFLYARRSEIRGVDIDPNYFNIFTAFTVPDIDDVAVIDFADSEER 1619
QY 173 -----FI-----174
Db 1620 LYWTDIKTOTITRAFINGTGLTVISRDIOQIRGLAVDWVRNLYWISSEFDETOINVAR 1679
QY 175 -----C-A-----W-----N-----PV--S 181
Db 1680 LDGSLKTSIIHGIDKPCQCLAAHPVRGKLYWTDGNTINMANMDGNSKILFQNKQEPVGLS 1739
QY 182 -D-----LL-V-SGSG-----D-----S-----TAR-I-----W--NL-- 198
Db 1740 IDYVENKLYWISSNGTINRCNLDOGNLEVIEMKEELTKATALITMDKKLWADONLAQ 1799
QY 199 -----S-----S-----E--NS-----TS-----204
Db 1800 LGTCNKRDRGNPSILNKTSIGVVMKVYDKEAQGSNSCOVNNGCSQCLPSETTRIC 1859
QY 205 -----G-----P-----206
Db 1860 MCTVGYVLQKNRMSQGISFLMYSVHEGIRGIRPLEPRDKVDALMPISGAFAVGDIFHA 1919
QY 207 -----T-----Q-----LV--L-R-----H-----C- 214
Db 1920 ENDTIYWTMDGLNKTSIRAKRQDTWKEDVVVNGLRVREGIADVIAWIAIGNIYWDHGFNLIEV 1979
QY 215 -----I-----R-----E-G-----GQ-----D-----V-- 222
Db 1980 ARLNGSFRVVISQGLDQPSRAVHEKGLFWTEWQVPCIGKARLDGSEKVMIVSVGI 2039
QY 223 --P--S-----NK-----D-----V--T-----S-LD-----W- 233

Db 2040 TWPNGSIDYEENKLYWCDAKSKIBRIDIDLTGANREVLGSGNVDLFSVAVFAGYIWS 2099
QY 234 -N-S- -E-GT- 238
Db 2100 DRAHANGSVRGHKNKDATETVWRTGLGVNLKIKIPNRVREKGTNVCAKENGCCOOLCL 2159
QY 239 -LLA- -T- -G- -S- 244
Db 2160 YRNSRRTCACAGYLADGVTCLRHGEGYLLYSGRITLKSILHSDETNLNSPVRPYENPN 2219
QY 245 Y- -D- -G- -F- -A- -R- -I- 251
Db 2220 YFKNIIALAFDYNQRREGTNRIFYSDAHFGNIQIIXDNMEDROVIVENVGSVEGLAYHRA 2279
QY 252 -W- -TK- -D- -G- -NL- 258
Db 2280 WDTLYWTSSTSSITRHTVDTQTRPGAIDREAVITMSDEDDHPHVLADDECONLMFTWNWNE 2339
QY 259 -A- -ST- -LG- -Q- -H- 265
Db 2340 QHPSIMRATLTGKNAHVVVSTDLTPNGLTIDHRAEKLIFSGLKIERCEVDGSQRHV 2399
QY 266 -K- -GP- -IF- -ALKW- -NK- -K- 277
Db 2400 IVKSGPGLFSLAVDYSYIFWS-DWGRRAILRSNKYTGTGETKILRSIDIPHQPMGIIVAN 2458
QY 278 - -G- -N- -F- - 280
Db 2459 DTNSCELSPCALLNGCHDLCLLTPGRVNCSCRGDRVLLANNRVTKNSSCNIYSEFEC 2518
QY 281 - -I- -L- -S- -AG- -V- 286
Db 2519 GNGDCVDYVLTCDGIPHCKDKSDEKLLYCNRSRSGFPKPCYNRCVPHGKCLDGTNDG 2578
QY 287 -D-K- -T- -I- 291
Db 2579 DSSDELDCVSTCSTVEFRCADGTCTIPRSARCQNQNMDCSDASDEKGCNNTDCTHFYKLGW 2638
QY 292 -I- -W- -D- -AH- -T- 297
Db 2639 KSTGFIRCNSLTVLPMSWICDGSNDGYSDELKCPVQNKHKCEYFGCPSGRCLNT 2698
QY 298 -G- -E- -A- -K- 301
Db 2699 WVCDGQKCEDGLDELHCSSCSWNQFACSVKCKISKHWICDGEDDCGSLDESISGA 2758
QY 302 -Q- -Q- -F- 304
Db 2759 VTCAADMFCQSGSHACVPQHWLDCGERDCPDGSDLSAGCAPNNTCDENAFCHNKVCI 2818
QY 305 P- -F- -H- -S- -AP- -ALD- -VD- -WQ- -S- 318
Db 2819 PKQFVCDHDDCGDGSDEFLQCYRQCQGPPEFRCA-DGRCLVNTLWQCDGDFDCPDSSDE 2877
QY 319 -N- -NT- -FASC- -STD- -MC- -I- -HV- -C- -KL 336
Db 2878 APINPRCRAEHSNCSFPM-CXNGRCIFPS-DGLCDIRDDCGDGSDETCHINECLSKKI 2935
QY 337 -G- -QD- -R- -PI- -KT- -F- -Q- -G- 347
Db 2936 SGCSQDCQDLPVSYKCKWPGFQKDDGKTCDVIDECSSGFPSCQOCINTYGYKCHCAE 2995
QY 348 -H- -T- -N- -E 351
Db 2996 GYETOPDNPNGCRSLSEBPFILADQHEIRKISTDGSNTYLLKQGLNNVIALDEFDYREE 3055
QY 352 -V- -NA- -I- -K- -W- -D- 358
Db 3056 FIYWIDSSRPNGRINRMCLNGSDIKVHINTAVPNALAVDWIGKLNLYWSDTEKRIIEVSK 3115
QY 359 -PT- -GNLL- -ASC- -S- 368

Db 3116 LNLGYPTVLVSKRLKFPRLDLSLDPAGNLYWIDCEPHYGRVGMGTNSQSVVIETKISR 3175
QY 369 -D- -D- -M- -TL- -K- -I- -W- 376
Db 3176 PMALTIIDYNNHRLYWADEHNIEFNSMDGSHRHKVPNQDIPGVIALTLFEDIYIWTGDKTK 3235
QY 377 -S- -MK- -Q- -D- -N- -C- -V 384
Db 3236 SLSRVHTSGADRLSLINSWHAITDIQVYHSYRQPDVSKHLCTVNGGCSHLCLLPGKTK 3295
QY 385 H- -D- -L- 387
Db 3296 HTCACTPNFYLAADNRITCLSNCTASQFRCKTDKICIPFMWKCDTVDDCGDSDPDCPEF 3355
QY 388 -Q- -A- -H- -NKB- -I- 394
Db 3356 KCQGRFQCGTGLCALPAFICDGENCGDNSDELNCDTHVCLAGQFKCTKNKKICIPVNL 3415
QY 395 -Y- -T- -I- -K- -W- -SP- -T 402
Db 3416 CNGQDDCGDEDEKDCPENSCSPDYFQCKTTHKICISKLAVCDDEDDPCADASDEANDCKKT 3475
QY 403 -GP- -G- -T- -N- -N- -P- 409
Db 3476 CGPHEFQCKNNNCIPDHWRCDNQDCSDNDEDNCKPOTCTTLKDFLCSNGDCVSSRFWCD 3535
QY 410 -N- -A- -N- 412
Db 3536 GBFDCAQDSEKNCETSCSKDOFQCSNGOCLSAKWKCDGHECKYGEDEKCEPAPFVCS 3595
QY 413 -LMLA- -SAS- -F- 420
Db 3596 SSEYMCASGGCLSLASKNGBPCDVGSDMDCVIECKEDQFQCKNKAYCIPIRWLCDGI 3655
QY 421 -D- -S- -T- -V- -R- -L- -W- -D- -V- 429
Db 3656 YDCVDGSDDETCGRGSGICRDDEFLCNSLCKLHFWVCDGEDDDCGDNSDEAPDMCVKFLC 3715
QY 430 -DR- -GI- -CI 435
Db 3716 PPTRPYRCNRDRIICLOLEKICNGINDCGNSDBEHCSGKLSKSPCKKDEFTCSNRNCI 3775
QY 436 -HT- 437
Db 3776 PMELQCDLDCGSDGDEOGLKTPIEHTCENNGNPGDDAYCQNIKTSVFCRCKPGFOR 3835
QY 438 -L- -T- -K- 440
Db 3836 NMKGRECADLNECLLFGICSHHCLNTRGSYKVCQDNFOEKNNSCIAKGSQALYIAND 3895
QY 441 -HO- -E- -P- -V- -Y- 446
Db 3896 TDILGFVPFNYSGGHQIISHVEHNSRITGMDVHYVORNVIWMSQPNPGGIFYKMIDARE 3955
QY 447 -S- -VA- -FS- 451
Db 3956 KQANSGLICPFKPRDIAVDVAGNVYVTDHSMHWFYSYTHHTWTSLSRYSINVQNLNG 4015
QY 452 -P- -DG- 454
Db 4016 PNCTRLLTNMAPEYAIANPKRGMWYTVIGDHSHEEAAAMDGTLLRVLVQKNLORPTG 4075
QY 455 -R- -Y- -LA- -S- -GS- 461
Db 4076 LTVDFHGERIYWADFELSIIIGSVLYDGSPPVSVSKQGLLPHRIDVFDYIYGAPKN 4135
QY 462 -F- -DK- -CV 466
Db 4136 GIFRVQKFGHGSVEVLALGVDTKSLVSHRYKQLNLPNPLDLSCDFLCLNPSGATCI 4195
QY 467 -H- -I- -W- -N- -T- 471
Db 4196 CPEGKYMNGTCHDDSLDSDCKLTCENGRCILNEKGLDRCHCWPSSYGGRCVNHCSN 4255

QY 472 --Q--T-G--AL--V--H--S--Y 480
 Db 4256 YCQNGGTCTPSTLGRPTCICALGFTGNCCKAVCEDSCHNGGCVVTAGNPYCHCQADY 4315
 QY 481 --R--G--T--G--GG--485
 Db 4316 TGRCOYVCHYVNSSECTIGNDGSVEVCPTRYEGPKCEIDKVCRCGHGHCINKDN 4375
 QY 486 --IF--E--VC--W--491
 Db 4376 EDIFCNCTNGKIASSQCLDGYCYNGGTGCLDPEISIPVCVCSTNWSGTQCPERPAPKSSK 4435
 QY 492 -----491
 Db 4436 SEHISTRSAIIVPLVLLVTLVTLVGLVWCKRRTKTIRROPIINGGINVEIGNPSY 4495
 QY 492 NA--AG--D-K--VG--A--SA--SD--G--505
 Db 4496 NMYEVHDHSDGGLLEPSPFIDPVKRYIGGSSAFKLPHTAPPYVINSDLKGLTFGPT 4555
 QY 506 --S--V--C-----VLDLRK 514
 Db 4556 NYSNPVYAKLYMDGQCRNSLASV-DERK 4583

RESULT 10

US-10-464-368-70
 ; Sequence 70, Application US/10464368
 ; Publication No. US20040023356A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Krumlauf, Robb
 ; APPLICANT: Ellies, Debra
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
 ; FILE REFERENCE: 40716-IP-017
 ; CURRENT APPLICATION NUMBER: US/10/464,368
 ; CURRENT FILING DATE: 2003-06-16
 ; PRIOR APPLICATION NUMBER: 60/388,970
 ; PRIOR FILING DATE: 2002-06-14
 ; NUMBER OF SEQ ID NOS: 140
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 70
 ; LENGTH: 4599
 ; TYPE: PRT
 ; ORGANISM: MOUSE
 US-10-464-368-70

Query Match 71.2%; Score 2612.9; DB 15; Length 4599;
 Best Local Similarity 9.5%; Pred. No. 6.5e-30;
 Matches 436; Conservative 55; Mismatches 17; Indels 4081; Gaps 381;

QY 1 MS-----I-----S-----4
 Db 1 MSQLLAILTLGLLPNAEVLIVGANQDHLCDPGEFLCHDHVTCVSQSLCDGDPDPCPD 60
 QY 5 SD-----EV-----N-----F-----10
 Db 61 QSDSLTCTPEVEIKPLNHIACHGSSACVHLSKLCNGVVDGPDGDEGHCOELLPC 120
 QY 11 --L--V--Y--R--Y-L--Q--17
 Db 121 QQLNCFKCAVRNATRCYEDGPEVAEDGSKQDECSIYIGCSQTKNTYGSYACSC 180
 QY 18 --SG-----F-----S-----H-----23
 Db 181 VE-GYMQSDNRCKVXHEPTDKAPMLLISSLETIELFYNGSKWTLSSANRNEIHTLD 239
 QY 24 -----S-----AF--T-----27
 Db 240 FIYSEMICWIESRESSNQLKCGQITKAGRLTDORIINSIQSFQNVQMAFDLITRNIFY 299
 QY 28 -----F-G-----IES--H-----33

Db 300 VDHVSDRI FVCNFGNSVCVTLLIESLHNPKAIAADPIAGKLFFTDYGVPKIERCDLDM 359
 QY 34 ---I---S---Q-----36
 Db 360 NRTRIVYKASQPSALALDLNRLVYVVDLYLDYVGVVDYQGNRHTIVQGRQVHLVGI 419
 QY 37 ---S---NI---NG-----A 42
 Db 420 TVFEDYLATSDNFNIIIRINFNGTDIHSIIKMESARGIRTYQKRTQPTVRSACEVDA 479
 QY 43 -----42
 Db 480 YGMPGGGSHICLLSSSYKTRICRRTGFNMGSDGSRCKPKNELFLFYKGRPGIVRGMD 539
 QY 43 L-----VP-----P-----S-II--OK--G-----L- 57
 Db 540 LNTKIADCEMPIENLVNPRALDFHAEANYIYFADTTSLIGRQKIDGTERETILKDDLD 599
 QY 58 --Q-----Y-----V--E-A-----E-----VS-IN-- 67
 Db 600 NYEGIAVDWIGNNLYWTDNGHKTINVARLEKASQSRKTLLEGGMSPRAIVVDPVNGWM 659
 QY 68 -----ED-----G-TL--P-----D-- 74
 Db 660 YWTDWKEDKIDDSVGRIBKAWMDGVNRQVFTSKMLWPNGLTLDFTSTLYWCDAIYDHI 719
 QY 75 ---G--R---P-----I-----E-SL-----81
 Db 720 EKVLNGTHRKVVYSGKELNHPFGLSHGNVYVFTDYMNGSIFQLDLMNVTLLRHERA 779
 QY 82 ---SL-I---D-----A--V-----87
 Db 780 PLFGLQIYDPRKQDQNMCRINNGCGTCLAI PAGRVACADNQLLDENGTTCTFNPEE 839
 QY 88 -----MP-----D-V-----V 92
 Db 840 IRFHICKPGEFRCKNKGICQARWKCDGDDCDLGDSEDSVTCFNHSCPDQKQNNRCI 899
 QY 93 -----QT--R--Q--Q-----AV--R-----DK- 102
 Db 900 PKRWLCDGANDCGSNEDESNOICTARTCOADQFSCGNGRCIPTAWLCLDREDDCGDQDEV 959
 QY 103 -----LAQ-----Q-----106
 Db 960 ASCERPTCEPLTQFICKSGRCISNKHWCDDDDCGDRSDEVCVHSCDLDQFRCSGRCI 1019
 QY 107 --H-A-----AAAA--A-----114
 Db 1020 PGHWACDGDNDGDFSEDETHINCKEARSAPAGCIGNEFQCRPDGNCIPDLMRCDGEKDC 1079
 QY 115 -----A-A--AT--N-----Q-----120
 Db 1080 EGSDEKNGNITRLCDHKTFCSTGRGINNAWVCDGDVDCEDSDDEDCDSFLCGPP 1139
 QY 121 -----Q-----GSAK-----NG-----127
 Db 1140 KYPCANDTSVCLQPEKLCNGRKCDPGSDEGLDCECLNNGGCSNHCSVVPGRGIVCSC 1199
 QY 128 --E-----N-T-----A--NGE-----E 135
 Db 1200 PEGHQLKDNRTCEIYDYCASHLRCSQVCEQKHVYKCSYEGWALGTGDESCTSVDSFE 1259
 QY 136 -----N-----136
 Db 1260 APTIFSRIRHRIIDLHKGYSLLVPLGRNTALDFHFNQSLLYWTDVVEDRIVRGKLS 1319
 QY 137 --G--A-----H-----T--IA-N-----N-----H 145
 Db 1320 SGGVSAIEVVVEHGLATPEGLTVDWIAGNIYWDNSLDQIEVSKLDSLRLATLIAGAMEH 1379
 QY 146 -----TD-----M-----M-----E-- 150
 Db 1380 PRAIALDPYRILFWTDWDANFPRISSASMGAGRTIYKDMKTGAWPNGLTVDHFERRI 1439


```
151 V-----DG-D-VEI-----P-----S-----N-----KA-----162
1440 VMTDARSDAISAFYDGTNMIIRGHEVLSHPFAVSLYSGEYVYWDRTINTLAKANKWT 1499
163 -----VV-----L-----RG-----H-----168
1500 GQNVSVIQTSAQPFQIYHPSQPAPNCPAANEGRGPCSHLCLINHNRSAAACPHL 1559
169 -----E-----SE-----V-----172
1560 MKLSSDKCYEMKKELLVARRSEIRGVODNPNVNFITFTVDDIDVAVIDFDASEER 1619
173 -----FI-----174
1620 LYWTDIKTQITRAFINGTLETVISRDIQSIIRGLAVDMVSRNLYWISSEFDEQTINVAR 1679
175 -----C-A-----W-----N-----PV--S 181
1680 LDGSLKTSIHGIDKPQCLAHPVRGKLYWTDGNTINMANMDGNSKILFQNOKEFVGLS 1739
182 -D-----LL-V-SGSG-----D-----S-----TAR-I-----W--NL-- 198
1740 IDIVENKLYWISSGNGTIIRCNLDGNGLEVIESKEELTKATALTMDKKLWADQNLAQ 1799
199 -----S-----E-----NS-----TS-----204
1800 LGTCNKRDGNPSILRNKTSIGVVHMKVYDKEAQQSGNSQVNVNGCGSQLCPTSETTRIC 1859
205 -----G-----P-----206
1860 MCTVGYLQKNRMSQGIIESFLMYSVHEGIRGIPLEPRDKVDALMPISGAAPAVGIDFHA 1919
207 -----T-----Q-----LV--L-R-----H-----C- 214
1920 ENDTIYDMGLNKISRAKRDQTKEDVVTNGLGRVEGIAVDWIAAGNIYWDTHGFNLIEV 1979
215 -----I-----R-----E-G-----GQ-----D-----V-- 222
1980 ARLNGSRFVVISQGLDQPSIAVHPKGFPLFTWENGQVPCIGKARLDGSEKVMIVSVGI 2039
223 --P--S-----NK-----D-----V-----T-----S-LD-----W- 233
2040 TWPNGISIDYEENKLYWCDAARSDKIERIDLDTCANREVLISGNSVDLFSVAVFGAYIWS 2099
234 -----N-S-----E-GT-----238
2100 DRAHANGSVRRGHKNDAETVTVMRTGLGVNLKEIKIFNRVREKGTNVCAKENGCCOOLCL 2159
239 -----LLA-----T-----G-----S-----244
2160 YRGNSRRTACAHGYLAGDGVTCRLRHEGYLLYSGRTILKSIHLSDETNLNSPVRPYENPN 2219
245 Y-----D-----G-----F--A-----R--I-----251
2220 YFKNIILAFDYNORREGTNRIFYSDAHFNGIQLIKDNWEDROVIVENVGSVEGLAYHRA 2279
252 -----W-----TK--D-----G-----NL-----258
2280 WDTLYTSSSTSSITRHTVQTRPGAIDREAVITMSDEDDHPHVALDDECONLMFWTNWNE 2339
259 -----A-----ST-----LG-----Q-H- 265
2340 QHPSIMRATLTGNAHVVSVDILTNGLTIDHRAEKLFPYSDGSLGKIERCEYDGSQRHV 2399
266 --K-GP-----IP-ALKW-----NK-----K-----277
2400 IVKSGFGTFLSAVDSYIFWS--DWGRAILRNSKVTGGETKILRSDIPHOPGIIAVAN 2458
278 -----G-----N-----F-- 280
2459 DTNSCBLSPCALLGGCHDLCLLTPDGRVNCSCRGDRVLLANNRCVTKNSSCNIYSEFEC 2518
281 -----I-----L-----S--AG-----V-----286
2519 GNGDCYDYVYLTCDDGIHPCKDKSDEKLLYCENRSCRGFKPCYNRRVCVPHGKLCODGTDCG 2578
287 -----D-K-----T-----T-I-----291
2579 DSSDELDCVKVSTCSTVEFRCADGTCIPRSARCNQNMDCSDASDEKGNNTDCTHIFYKLG 2638
292 -----I-----W-----D-----AH-----T 297
2639 KSTGFRRCNLSLCLVLPSPMICDSNDCGDYSDDELKCPVQNKHKCEENYFCPSGRCLINT 2698
298 -----G-E-----A-----K-----301
2699 WVCDSGQKCEDGIDELHCDSSCSWNQFACSVKVCISKHMICDGEDDCGDSLSDSDSICGA 2758
302 -----Q-----Q-----F-----304
2759 VTCADMFSCQSHACVPOHMLCDGERDCPDGSDLSAGCAPNNTCENAFMCHNKVCI 2818
305 P--F--H-----S-----AP-----ALD-----VD--WQ-----S-- 318
2819 PKQFVCDHDDDCGSDGSEFLQCGYRQCGPEEPRCA--DGRCLVNTLWQCDGDFDCPDSSDE 2877
319 --N-----NT--FASC-----STD-MC-I-----HV--C--KL 336
2878 APINPRCRAEHSNCSFFEM--CKNGRCIPS--DGLCDIRDGCGSDGTNCHINECLSKKI 2935
337 -G-QD-R-PI-----KT-----F-----Q-----G-----347
2936 SGCSQDCQDLPSYKCKCWPGFQKDDGKTCVDIDECSSGFPSCSQCCINTYGYKCHCAE 2995
348 -----H-----H-----T-----N-----E 351
2996 GYETQPDNPGCNSLSDDEPFLILADQHEIRKISTDGSNTLLKQGLNVNIALDFDYREE 3055
352 -----V-NA-----I-K--W-D-----358
3056 FIYWDSSRPNGSRINRMCLNGSDIKVHVHTAVPNALAVDWIGKLYMSDTEKRIIEVSK 3115
359 -----PT-----GNLL-ASC-----S- 368
3116 LINGLYTILVSKRLKFPRLDLSLDPAGNLYWIDCCBYPHIGRVMGDMGTNSVVIETKISR 3175
369 -----D-----D-----M-----TL-K-I-W-----376
3176 PHALITIDYNHRLYWADENHIEFSNMDGSHRHKVPNQDIPGVIALTLFEDYIYWDGKTK 3235
377 --S--MK-----Q-D-----N-----C-----V 384
3236 SLRSHVKTSGADRSLSLNSWHAITDIQVYHSVRQPDVSKHLCTVNVNGGCSHLCLLGPCKT 3295
385 H-----D-----L-----387
3296 HTACPTNFYLAADNRNLTCLSNCTASQFRCKTDKCIPIFWMKCDTVDDCGSDGSDPDCEPF 3355
388 -----Q-----A-----H-----NKE-I-----394
3356 KCQPGRFQGTGLCALPAFICDGENCGNSDELNCDTHVCLAGQFKCTKNKKCIPVNL 3415
395 -----Y-----T-----I-K-W--SP-----T 402
3416 CNGQDDCGDEDEKDCPENSCTPDYFQCKTTHKICISKLWVCDDEDDPCADASDEANCCKKT 3475
403 -GP-----G-----T-N-----N-P-----409
3476 CGPHEFCQKNNCIPDHWRCNDQNSDSDENCKPQTCCTLKDFLCSNGDCVSSRFWCD 3535
410 -----N-----A-----N-----412
3536 GBFPDCADGSEKNCETSCSKDQFQCSNGQCLSAKWCDGHEDCKYGEDEKNCBPAPVCS 3595
413 --LMLA-----SAS-----F-----420
```

Db 3596 SSEYMCASGCLSLASLKCNGEPCDVGSDMDVIECKEDQFQCKNKNKAYCIPRWLDCGI 3655
QY 421 ---D-S---T---V-R---L-W---D-V---429
Db 3656 YDCVDSDEETCGRGSGICRDEDFLNCNLSCKLHFWVCDGEDDCGSDNSDEAPDMCVKFLC 3715
QY 430 ---DR---GI---CI 435
Db 3716 PPTPRYCRNDRIQLQLEKICNGINDCGDSDBEHCGKLSKSPCKCKDEFTCSNRNCI 3775
QY 436 ---HT---437
Db 3776 PMELQCDSLDCGSDGDEQCLKTPTEHTCENNGNPGCDDAYCNQIKTSVFCRCKPGFOR 3835
QY 438 ---L-T---K---440
Db 3836 NMKGRECADLNECLLFGICSHHCLNTRGSYKVCQDNQFQEKNNCSIAKSGEDQALYIAND 3895
QY 441 ---HQ---E---P-V-Y---446
Db 3896 TDILGFVYPNYSGGHQISHVEHNSRITGMVHYQRNVIVWSTQFNPGGIFYRMIDARE 3955
QY 447 ---S---VA---FS---451
Db 3956 KQANGSLICPEFKRPDIADVWAGNVYWDHSMHWFYSYTHWTSLSYSINVGQNG 4015
QY 452 ---P---DG---454
Db 4016 PNCTRLTNMAGBPYATVNPKRGMYWTVIGDHSHEEAMGTLRLRVLQKNLQRTG 4075
QY 455 ---R-Y---LA---S---GS---461
Db 4076 LTVDFGERIYADFELSIITGSVLVDGSPVSSKQGLLPHRIDVPEDYIYGAPKN 4135
QY 462 ---F---DK---CV 466
Db 4136 GIFRVQFGHGSVEVLALGVDTKTSILVSHRYQLNLPNCLDLSCDFCLINPSTGATCI 4195
QY 467 ---H---I---W---N-T- 471
Db 4196 CPEGKYMNGTCHDDSLDDSKLTCENGRGCIINEKGLRCHCWPSYSGRCVNHCSN 4255
QY 472 ---O---T-G---AL---V---H---S---Y 480
Db 4256 YCQNGGTCTPSTLGRPTCICALGFTGPNCGKAVCEDSCHNGSCVVTAGNPQYCHCOADY 4315
QY 481 ---R---G---T---GG---485
Db 4316 TGRQCQYVCHYCVNSECTIGNDGSVEVCPTRYEGPKCEIDKVCVCHGHCINKN 4375
QY 486 ---IF---E---VC---W---491
Db 4376 EDIFCNCTNGKIASQCQCDGVCYNGGTCQLDPETSPVVCVTNWSGTQCPERPAPKSSK 4435
QY 492 ---NA---AG---D-K---VG---A---SA---SD---G---505
Db 4496 NMYEVDHSDGGLLEPSFMDIPVKSRYIGGSSAFKLPHTAPPYILNSDLKGLPTFGPT 4555
QY 506 ---S-V---C---VLDLRK 514
Db 4556 NYSNPVYAKLYMDQNCNRSLSV-DERK 4583

RESULT 11

US-10-120-801-74

; Sequence 74, Application US/10120801

; Publication No. US20030203843A1

; GENERAL INFORMATION:

; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Sytek, Kimberly
; APPLICANT: Mehraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 74
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-120-801-74

Query Match 71.2%; Score 2612.1; DB 14; Length 5175;
Best Local Similarity 9.2%; Pred. No. 1.2e-29;
Matches 446; Conservative 53; Mismatches 12; Indels 4339; Gaps 386;
QY 1 M-----SI-----SS-----D-----E-----7
Db 1 MGRSPSWLYGVGLGILLATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLQVREGAAK 60
QY 8 ---V---N---F---L-VY-----13
Db 61 IFKTVMAQREKLIYNIMVPHDPYLGEIINTDSTYFMQLSKVYVHGGDCPEKLTG 120
QY 14 ---R---Y---L---QF---SG---F-----SHSAF-T 27
Db 121 ILKALQISLPSSFIYVFTDARSKDYLHLEDEVLATIQEKSSVVFVMTGDCGNRTHPGPRT 180
QY 28 ---FG-I---E---S---H-ISQ-----S-NI-----39
Db 181 YEKIAAASFGQVPHLEKSDVSTVLEYVRHAVKQKVHLMYAEARGGTGVSGRNIPVDKHL 240
QY 40 ---N---N---GA-----LV---P-P-----46
Db 241 ELTISLSDKDDSDNLDIVLRDPGRTVDKRLYSKEGGTIDLKNVKLIRLKDPSPGVTV 300
QY 47 ---A-----A-----LI---S-I-----52
Db 301 NTNSRLKHTIRVFGHGAVDPKYGFASRPLDRIELARPPVNLQDTYLLINMTGLIPGTV 360
QY 53 --I-----Q-----KGL-----O-Y-----VE-----AE- 63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPVFKGLFFVRVQGVQDDEDEPMRIAPT 420
QY 64 ---V-----S-I-----N-----ED-----GTILF-----73

Db	421	AIGSVTVGGPRAFMSP	IHOEFVGRDLNLSCTVESASAYTIYWKGTGDIIGPLFYHNTD	480	Db	1499	DKTSWKTSDDKRKLHVFKAKITD	SGYKVCVARNAGEGSKSFQVEVIVPLNLDSEKYYKKK	1558													
Qy	74	-----D-----	-----G-----	-----R-----P-----I-----	-----E-----	79	Qy	217	-----EG-----	-----G-----Q-----D-----V-----P-----	-----SN-----	225										
Db	481	TSWVTIPELSLKDAGEYECRVISNNGNSYKTRVETRESPPPEIFGVNRVSVPLGEAFLH	540	Db	1559	VPAKEGEEVTLGCPVSGPVPPOINWVVDGTVPVPGKKYKATLSNDGLTLHPDSVSUKQE	1618	Qy	226	-----S-----	-----L-----	-----D-----	-----W-----	233								
Qy	80	-----S-----	-----L-----	-----SL-----I-----	-----DA-----	86	Db	1619	GNHCVASQKGNILIDVELSVLAVPIVGEDDNLEVLFGKDISLSCLQTESDDKTTFW	1678	Qy	234	-----N-----SE-----	-----G-----	-----TL-----	-----L-----	240					
Db	541	CSTRSAGEVEIRWTRYGATVFGNPNTRNPTNGTLKIHVTRADAGVYECMARNAGMST	600	Qy	87	-----VM-----	-----P-----DV-----	-----V-----	92	Db	1679	SINGESDRPDNDVQIPSGHRLYITDAEPENNGKVMCRVTNSAGKAERTLTLDVLEPPVF	1738	Qy	241	-----A-----	-----TG-----	-----SY-----	-----D-----	246		
Qy	93	Q-----	-----T-----R-----Q-----OAYRD-----	-----KLA-----	-----Q-----	105	Db	1739	VEPVFEANQKLGNNPIILQCOVTGNPKPTVIWKIDGNDVKSFLFDESLSLLRIEKLGT	1798	Qy	247	-----G-----	-----F-----	-----A-----	-----R-----	-----I-----	251				
Db	661	QDSKFLYIRDATHDEGTTECRAMSAGQA-RDTTDLMLATPPKVEIIONKMMVGRGDRV	719	Qy	106	-----	-----Q-----H-----A-----	-----AAA-----	113	Db	1799	KSAQISCTAENKAGTASRDFFIQNTAAPTFKNEGQETIFRESSETITLDCPVSLGDFQIT	1858	Qy	252	-----W-----	-----T-----	-----KD-----	-----	255		
Qy	720	SPECKTIRKPHKIRWFKNGKDLIKPDDYIKINEGQLHMGAKDEADAGAYSCVGENMAG	779	Db	114	-----AAAA-----	-----T-----N-----	-----Q-----	GS-----A-----KN-----	126	Qy	1859	WMKQGLPLTENDAIFTLNTRLTILNANRDHEDIYTCVANNTAGQVSKDFVWVQVLPKI	1918	Qy	256	-----	-----GN-----	-----L-----	-----A-----	-----S-----	260
Db	780	KDVQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIIEWQKGNVLLATLNNP	839	Qy	133	-----G-----	-----EE-----	-----N-----	-----G-----	137	Db	1919	KNAVTLNEEGEELITCDAGNPTPTAKWDFNQDLPKEAVFVNNHTVVVNNVTKYH	1978	Qy	261	-----T-----	-----L-----	-----GQ-----	-----H-----	-----K-----	266
Qy	840	RYTLADG-NLLITDAQIEDQOFTCIARTYVQSQSOTLMTGLVSPVLGHVPPEBOL	898	Db	138	-----A-----	-----H-----	-----TI-----	-----A-----N-----NH-----	145	Qy	1979	TGVYKCATNKVGQAVKTIINVHRTKPRFESGLTESELTWNLTSTRITLDCVDVDAIGVGI	2038	Qy	267	-----G-----P-----	-----I-----	-----F-----	-----	270	
Qy	899	IEQDLTLSCVVVLGTPKPSIVIKDDKPVEGPTIKIEGGSLRLRGNPKDEKGYTC	958	Qy	146	-----	-----	-----T-----	-----	146	Db	2039	SWTVNGKPLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTNLFVQVP	2098	Qy	271	-----	-----AL-----	-----K-----	-----W-----	-----	274
Qy	138	-----A-----	-----H-----	-----TI-----	-----A-----N-----NH-----	145	Db	2099	PTIVNEGGETVIENNSVLPCVETGKPNPVVTWKDGRPVGLKSVQVLSQVQKPIVH	2158	Qy	275	-----	-----N-----	-----K-----	-----KG-----	-----	278				
Qy	959	IAPVAGNGLTHINVQLIKKPEVPYKPEGIVFKPTISGMDEKHVAVNSVTHDVLDEGF	1018	Qy	1079	DNEQKTIIRMTPIMLSPGSSFNWVDDLFTIPCDVYGDPRPVITWLLDDKPFTEGVN	1138	Qy	2159	ABIAHKSIVYKAKNDVGTAEISFDVDIITRPMIOKGIKNIITAKGKALPKFCPIDDDK	2218	Qy	279	-----NF-----	-----IL-----	-----S-----	-----AG-----	285				
Qy	1139	EDGSLTIPVNEAHRGFTTCHAQNAAGNDRTRVTLTVHTTPTINAENQEKIALQNDIVL	1198	Qy	156	-----I-----	-----PS-----	-----N-----K-----A-----	-----VVL-----	165	Qy	2219	NFKGOIILWRNYQPIDLEADARITRLSNDRRLTLNVTENDEGGQYSCRVKNDAGENSFD	2278	Qy	286	-----V-----	-----D-----KT-----	-----TI-----	-----I-----W-----D-----A-----H-----	296	
Qy	166	-----	-----R-----	-----	-----	167	Qy	2279	FRATVLPPTIIMLDKDKKTAVEHSTVTLSCPATGKPEPDITWFKDGAHIENTADI	2338	Qy	297	-----T-----	-----G-----	-----EA-----	-----K-----	-----Q-----	302				
Qy	1199	ECPAKALPPPVLRLWYEGEKIDSLIPHITREDGALVLQNVKLENTGVFVQVSNLAGED	1258	Qy	168	-----HE-----	-----SEV-----	-----F-----I-----C-----A-----	-----WNP-----V-----	180	Qy	2339	PNGELNGQLKTRIKEGDAGKYTCADNSAGSVEQDVNVVNTIPIKIEKOGIPSDYESQ	2398	Qy	303	-----Q-----	-----	-----	-----	303	
Qy	1259	SLSYTLTVHEKPIIIEVPGVDVVKGTIEIPCRATGVEVIRTNWKNKIDLMKDEKPF	1318	Qy	181	S-D-L-L-V-S-----	-----G-----	-----SG-----	-----D-----	193	Qy	2399	QNERVVISCFVYARPPAKITWLKAGKPLQSDKFVKTSANGQKLYLPKLRETOSSKYTCIA	2458	Qy	304	-----F-----	-----P-----F-----	-----H-----	-----S-----	308	
Qy	1319	SVDNLGTLRIYEADKNDIGNYCNVVTNEAGTSQMTTHVDVQEBPIILPSTQTNNTAVGD	1378	Qy	194	RI-----	-----W-----	-----NL-----	-----S-----ENST-----	204	Qy	2459	TNEAGTKRDFKVKMLVAPSPDEPNIVRITVNSGNPSTLHCPAKSPFTITWLKDGNA	2518	Qy	309	-----	-----A-----	-----P-----	-----	310	
Qy	1379	RVELKCYEASPPASVTFRRGIAIGTDTKGYVVEDGTLVLIQOSASVEDATIYTKASNP	1438	Qy	205	-----G-----	-----P-----	-----TO-----	-----L-----V-----	211	Qy	2519	IEPNDRYVFDAGRQLQISKTEGSDQGRYTCIATNSVSGDDELNTLEVIIPPIDGERRE	2578	Qy	311	-----A-----	-----L-----D-----	-----VD-----	-----WQ-----	-----S-----N-----	319
Qy	1439	AGKAENLQVTVIASDDIKDDPVVTOESIKESHPSFLYCPVFSNPLPQISWYLNKPLID	1498	Qy	212	-----R-----H-----	-----	-----CI-----R-----	-----	216	Qy	2579	AVAVIEGFSSELFCDSNSTGVDEWQKGLTINQDTRLRGSFQIPSSGKKMSFLSARKS	2638								

QY 320 -----N---T-----FA-----S-----C---S---T--- 327
Db 2639 DSGRYTCIVRNPAEARKLDFEAVNDPPSISDELSANIQTIVPYVPEINCVSQSPHP 2698
QY 328 -----DM-----C-----I--- 331
Db 2699 KVVWLPDDKLEPDSAAYELTNNGETLKIVRSQVEHAGTYTCEAQNNGVKARKDFLVRVT 2758
QY 332 -----H-----V-----C-----KL-----G----- 337
Db 2759 APHFKEKEBEVAVRGDTMLLTCAESSVPLSVYWHADHESVQNGVITSKYAANEKTL 2818
QY 338 -----Q-----Q-----C-----KL-----G----- 343
Db 2819 NVTNIQLDDEGYCTAVNEAGITKFKFLIVITIETPYFLDQKLYPIILGKRLTLDCSAT 2878
QY 344 -----T---F-----N-----E-----VN-----A----- 345
Db 2879 GTPPTILFMKDKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSEKDMWVE 2938
QY 346 -----Q-----Q-----G---H----- 348
Db 2939 VLLPPKLSKEWINVEVQAGDPLTECPIEDTSGVHITWSRQFGKGQOLDMRAQSSDKSK 2998
QY 349 -----T-----N-----E-----VN-----A----- 354
Db 2999 LYIMQATPEADSYCIANDAGABAVFOVTNTPPKIFGDSFSTEIVADTTLIPCR 3058
QY 355 -----IKW-----P-----T-----GN----- 358
Db 3059 TEGIPPEISWFLDGKPILEMPGVYKQGLSLRIDNIKENQGRYTCVAENKAGRAEQD 3118
QY 359 -----P-----T-----GN----- 362
Db 3119 TYVEISEPPRVVMASEVMRVEGQATTIRCEVEGNEPVPVWNLKDGEPYTSLLQFSTKL 3178
QY 363 -----L---LA-----SC-----S---D---D-----D-----M--- 371
Db 3179 SYLHLRETTLADGGTYTCIATNKAGESQTTTDEVILVPPRIEDEERVLOGKEGNTYMHVC 3238
QY 372 -----L-----S-----D-----T-----L----- 373
Db 3239 QVTGRPVVYTKRNGKEIQFNPVLHIRNATRDEKYSKIASNEAGTAVADFLIDVFT 3298
QY 374 -----KI-----W-----S-----M---K 379
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGGRPFNMDNIILSPRGDTLMILK 3358
QY 380 -----QD---N-----C----- 383
Db 3359 AQRFDGGLYTCVATNSYGDSEQDFKNVYTKPYIDETIDOTPKAVAGGEIILKCPVLGNP 3418
QY 384 -----Y---H-----DL-----Q-----A-----H---NKE-I 394
Db 3419 TPTVTKRGDDAVPNDNRHTIVNNYDLKINSVTTEDAGQVSCIAVNEAGNLTHYAAEVI 3478
QY 395 -----Y---TI---K-----W-----SP----- 401
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLFSISWFRGDKFVYLYDRYSISFPGSHIT 3538
QY 402 -----T----- 402
Db 3539 INKAKLSDGKYTCRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPI 3598
QY 403 -----G---P-----T-----NN----- 408
Db 3599 SGIPQPDVITKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRVTCIATNRRGKASHD 3658
QY 409 -----P----- 409
Db 3659 FSLDVLSPPEFDIHGTQPTIKREGDTITLTCPIKLAEDTADQVMDVSWTKOSRALDGDLT 3718

QY 410 -----N-----A---N---L-----ML---A---SAS---F-----D----- 421
Db 3719 DNVDISDDGRKLTISOASLEENAGLTCTIALNRAGEASLEFKVEILSPPIDISRNDVQPO 3778
QY 422 -----S-----T-----VR-----L-----W----- 427
Db 3779 VAVNQPTIMRCVATGHPFPSIKWLKNGKEVTDENIRIVEQGOVQLILRTSDSHAGKWSC 3838
QY 428 -----DV-----D---R---G---I---C----- 434
Db 3839 VAENDAGVKELEWLVDFTPPVVSVKSDNPRIKALGETITLFCNASGNPYPQLKWKAGGSL 3898
QY 435 -----I---H-----T----- 437
Db 3899 IFDSPGARIKSGARLDIPLHKKTVDGYTQCALNAAGTSEASVVDVLVPPEINRDGI 3958
QY 438 -----LT-----K-----H-----Q----- 442
Db 3959 DMSPLPAQOGLTLOCLAQKPVQPMRWTLNGTALTHTPGITVASDSTFIQINNVSLS 4018
QY 443 -----E-----PV-----Y-----S--- 447
Db 4019 KGVTCTAENVAGSDNLMYNDVWQAPVISNGGTQKQVIEGELAVIECLVEGYPAPQVSWL 4078
QY 448 -----V-----A----- 449
Db 4079 RGNRVETGVQVRYVTDGRMLTIIEARSLDSGIYLCSATNEAGSAQOAYTLEVLVSPKI 4138
QY 450 -----FS-----P-----DG----- 454
Db 4139 ITSTPGVLTSPSSGKSFPCAARGVYDPPIISWTLNGNDIKDGENGTIGADGTLHIEAE 4198
QY 455 -----R---Y-----L-----A---SG----- 460
Db 4199 ERHLIYECTAKNDAGADTLEFPVQTIIVAPKISTSGNRYINGSEGTETVIKCEISESEF 4258
QY 461 -----S---F---D---K-----C----- 465
Db 4259 SMSKNGVPLLPNNLIFSEDYKLIKILSTRSDQGEYSCTAANKAGNATQKTNLVGVAP 4318
QY 466 -----VH-----I---W-----NT----- 471
Db 4319 KIMERPTQVHKGDQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTKKSVIFSS 4378
QY 472 -----O-----TG----- 474
Db 4379 ISPSQAGVYTCKAENWVASTEEDIDLIVMPEVWPERMNVSTNPRQTVFLSCNATGIPE 4438
QY 475 -----A-----L---V-----H----- 478
Db 4439 PVIWNRDSNIAIONNEKYQILGTTLAIERNVLPDDGPFYHCIAKSDAGQKIATRKLVNK 4498
QY 479 -----S---YR---G--- 482
Db 4499 PSDRPAPIWVECEDEKPKKTEYMDIRGDPDDNPOLLPKWQVEDSSLSNGSIAYRCMPGP 4558
QY 483 -----T-----G---I---FE--- 488
Db 4559 RSRVTLLHAAPQFIVKPKNTAAIGAIIVELRCSAAGPPHPTITWAKDGKLIEDSKFEIA 4618
QY 489 -----V---C--- 490
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKSSNQKNVAVITCYE 4678
QY 491 -----W-----N---A----- 493
Db 4679 RNOQVSRGLTWEYNGVPMKPNLAGIHFMMNGSLIVLDTSSLKEGDLLEYTCVKVNRHRS 4738
QY 494 -----A---G---DKV---G---AS--- 502
Db 4739 IPHLTSAFEGPEVKTIDIKVNNVNGSVVLDCVETSDPLTTHVVVTKNQKMLDDDAIYV 4798
QY 503 -----S-----D---GSV---CV-----LD---L-----R---K 514

Db	4799	LPNNSVLLNVEYDEG-VYKVASNSIGKAFDDTQLNVTYGGSSREAYK	4847	Db	601	RKMRLDIMEPPSVKVTQDVYFNMREGVNLSCAMGDPKPEVHWPKGRHLLNDYKYQVG	660
RESULT 12							
US-10-369-493-6859							
; Sequence 6859, Application US/10369493							
; Publication No. US20030233675A1							
; GENERAL INFORMATION:							
; APPLICANT: Cao, Yongwei							
; APPLICANT: Hinkle, Gregory J.							
; APPLICANT: Slater, Steven C.							
; APPLICANT: Goldman, Barry S.							
; APPLICANT: Chen, Xianfeng							
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF							
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES							
; FILE REFERENCE: 38-10(52052)B							
; CURRENT APPLICATION NUMBER: US/10/369,493							
; CURRENT FILING DATE: 2003-02-28							
; PRIOR APPLICATION NUMBER: US 60/360,039							
; PRIOR FILING DATE: 2002-02-21							
; NUMBER OF SEQ ID NOS: 47374							
; SEQ ID NO 6859							
; LENGTH: 5175							
; TYPE: PRT							
; ORGANISM: Caenorhabditis elegans							
US-10-369-493-6859							
Query Match							
Best Local Similarity 9.2%; Pred. No. 1.2e-29;							
Matches 446; Conservative 53; Mismatches 12; Indels 4339; Gaps 386;							
QY	1	M-----SI-----SS-----D-----E-----7		QY	1019	AIPCVSVGTPPIITWYLDGRPIITNSRDTVTADNTLIVRKADKSYGVYTCOATNSAG	1078
Db	1	MGRSPWLYGVLLGALLATTCSSVNDKNDPTCKSLAFVFDITGSMFDDLVQVREGAAK	60	QY	147	D-----M-----E-----VD-----GD-----V-----155	
QY	8	-----V-----N-----F-----L-----VY-----13		Db	1079	DNEQKTIIRIMTPIISMISPCQSSFNWVDDLFTIPCDYGDPKPVIWLLDDKPFTEGVN	1138
Db	61	IPKTVMAQREKLIYNIWPFHPDYLGEIINTDSTYFMRLSKVYVHGSGDCPEKTLTG	120	QY	156	E-----I-----PS-----N-----K-----A-----VVL	165
QY	14	-----R-----Y-----L-----QE-----SG-----F-----SHSAF-T	27	Db	1139	EDGSLTIPNVNEAHRGFTTCHQAAGNDRTRVTLTVHTTPTTINAEQKIALQNDIVL	1198
Db	121	ILKALQISLPSSFIYFTDARSXDYHLEDEVLTIOEKSSVVFVMTGCGNRTHPGFRT	180	QY	166	-----R-----G-----167	
QY	28	-----FG-I-----E-----S-----H-----ISO-----S-----NI-----39		Db	1199	ECPAKALPPVRLWTEYGEKIDSLIPHTREDGALVLQNVKLENTGVFCVQVSNLAGE	1258
Db	181	YEKIAASFQVPHLEKSDVSTVLEVRHAKVKVHLYEARERGGTVSRNIPVDKHL	240	QY	168	-----HE-----SEV-----F-----I-----C-----A-----WNP-----V-----180	
QY	40	-----N-----GA-----LV-----P-----P-----46		Db	1259	SLSYTLTVHEKPKIIEVPGVVDVVKGFTEIEPCRATGYPEVIRTWNKCIDLKMDEK	1318
Db	241	ELTISLGDKDDNDLIVLRDEGRVTDKRLYSKEGGTIDLKNVKILRLKOPSPGVWTV	300	QY	181	S-D-L-L-V-S-----G-----SG-----D-----S-----TA-----193	
QY	47	-----A-----A-----LI-----S-----I-----52		Db	1319	SVDNLTGLRIYEADKNDIGNVNCVVTNEAGTSQMTTHVDVQEPPIILPSTQNTNNTAVGD	1378
Db	301	NTNSRLKHTIRVFGHGAVDKYGAFASRLDRIELARPRVNLQDTYLLINMTGLIPPGTV	360	QY	194	RI-----W-----NL-----S-----ENST-----S-----204	
QY	53	-----I-----Q-----KGL-----O-----Y-----VE-----AE-----63		Db	1379	RVELKCYVEASPPASVTWFRFGIAGTDTKGYVVEDSGTLVIOQASVEDATIYTCKASN	1438
Db	361	GEIDLVDYHGSHLYKAVASPHRTNPNMYFAGPPVPPKGLFFVRVQGYDSDNVEFMRIAPT	420	QY	205	-----G-----P-----TO-----L-----V-----L-----211	
QY	64	-----V-----S-----I-----N-----ED-----GTLF-----73		Db	1439	AKAEANLQVTVIASPDIDKDPDVVTOESIKESHPPSLYCPVFNPLPQISWYLNDRPLD	1498
Db	421	AIGSVVGGPRAPMSPHQEFVGRDLNLSCTVESASAYTIYVKTGEDIIGGLFFHYHTD	480	QY	212	-----R-----H-----CI-----R-----216	
QY	74	-----D-----G-----R-----P-----I-----E-----79		Db	1499	DKTSWKTSDDKRLHVFKAITDSGVYKCVARNAGEGSKSFQVEIVPLNLDSEKYYKK	1558
Db	481	TSVWTIFELSKDAGEYECRVISNNNGYVKTRETRESPPPEIFGVNRVSVPLGEAFLH	540	QY	217	-----EG-----G-----Q-----D-----V-----P-----SN-----225	
QY	80	-----S-----L-----I-----DA-----86		Db	1559	VFAKEGEVTLGCPVSGFPVQINWVVDGTVEVPGKKYKATGLNSDGLTLHDFSVSVKQE	1618
Db	541	CSTRSAGEVEIRTRYGATVNGPNTERNPTNGTLKIHVTRADAGVYECMARNAGMST	600	QY	226	-----KDV-----S-----L-----D-----W-----233	
QY	87	-----VM-----P-----DV-----V-----92		Db	1619	GNVHCVAQSGKNLIDIDVELSVLNPVIGEDDNLEVLFGKDLSLSCDLQTESDKTTFFW	1678
				QY	234	-----N-----SE-----TL-----L-----240	

Db 1679 SINGSEDRPDVQIPSDGRLIYITAKPENNGKYMCRVNTNSAGKAERTILTLDVLEPPVF 1738
QY 241 -A- -TG- -SY--D- 246
Db 1739 VBPVFEANOKLIGNPPILOQCVTGNPKPTVIWKIDGNDVDRKSWLFDLSLLRIEKLATG 1798
QY 247 -G- -P- -A- -R- 251
Db 1799 KSAQISCTAENKAGTASRDPFIQNIAPTAKNEGQETIFRESEITITLDCPVSLGDFQIT 1858
QY 252 W- -T- -KD- 255
Db 1859 WKQGLPLTENDAIFILDNTRLTILNANRDHEDIYTCVANNTAGQVSKDFVVQVLPXI 1918
QY 256 -GN- -L- -A- -S- 260
Db 1919 KNAVTLTLENEGEIILTCAEGNPTPTAKWDFNQGDLPKEAFVANNHTVVVNNVTKYH 1978
QY 261 T- -L- -Q- -H- -K- 266
Db 1979 TGVKCYATNKVQAVKNTINHVRTKPRFESGLTESELTVNLTRSLTECDVDDAIGVGI 2038
QY 267 -G-P- -I- -F- 270
Db 2039 SWTNGKPFPLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTNLFVQVP 2098
QY 271 -AL- -K- -W- 274
Db 2099 PTIVNEGGEYTVIENNSLVPCEVTGKPNPVVWTGDRPVGDLKSVQLSEGOQFKIVH 2158
QY 275 -N- -K- -KG- 278
Db 2159 AEIAHKSYSICMAKNVDGTAETISFDVDIITRPMIOGKINIVTAIKGALPFPKCPIDDDK 2218
QY 279 NF- -IL- -S- -AG- 285
Db 2219 NFKQIILWLNYPIDLEAEDARITRLSNDRRLTILNVNTEDEGOYSCRVKNDAGENSFD 2278
QY 286 -V- -D- -KT- -TI- -I-W--D--A-H- 296
Db 2279 FKATVLVPPTIIMLDKDKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAIIHENIADI 2338
QY 297 -T- -G- -EA- -K- -Q- 302
Db 2339 PNGELNGNQLKITRIKEGDAGKYTCESADNSAGSVEQDVNVNVTIPKIEKDGIPSDYESQ 2398
QY 303 Q- 303
Db 2399 QNERVWISCPVYARPPAKITWLKAGPLQSDKFVKTSANGQKLYLFKLRETDSSKYTCIA 2458
QY 304 -F- -P- -H- -S- 308
Db 2459 TNEAGTKDRDPKVSMLVAPSFDEPNIVRRITVNSGNPSTLHCPAKGSPSPTITWLKDGNA 2518
QY 309 -A- -P- 310
Db 2519 IEPNDRVFFDAGQLQISKEGSDQGRYTCIATNSVGSDDLENTLEVIIPPIDGERRE 2578
QY 311 -A- -L- -D- -VD- -WQ- -S-N- 319
Db 2579 AVAVIEGFSSSELCFDSNSTGVVQKDLTINQDTLGRDSFIQIPSSGKKMSFSLARKS 2638
QY 320 -N- -T- -FA- -S- -C- -S- -T- 327
Db 2639 DSGRYTCIVRNPAGEARKLDFAVNDPPSISDELSSANIQTIVPYVPEINCVVSGSPHP 2698
QY 328 -DM- -C- -I- 331
Db 2699 KYVWLFDDKPLEPDSAAVELTNNGETLKI VRSQVEHAGTYTCEAQNNVGKARKDFLRVT 2758
QY 332 -H- -V- -C- -KL- -G- 337
Db 2759 APPFEKEREBEVARVGDMLLTCAESSVFLSSVYVWHAHDESQVNGVITSKYAANEKTL 2818

QY 338 -Q- -DR- -PI- -K- 343
Db 2819 NVTNIQLDDEGYCYTAVNEAGITKKFKLIVETPYFLDQOKLPIILGKRLTLDCSAT 2878
QY 344 -T- -F- 345
Db 2879 GTPPTILLPMKDGKRLNESDEVDIIGSTLVIDNPQKEVGRYTCIAENKAGRSEKDMMVE 2938
QY 346 -Q- -G- -H- 348
Db 2939 VLLPPKLSKEMINVEQAGDPLTLECPIDBTSGVHITWSRQFGKQGLDMRAQSSDSK 2998
QY 349 -T- -N- -E- -VN- -A- 354
Db 2999 LYIMQATPEDADSYSCIAVNDAGGAFAVQVTVNTPPKIFGDSFSFTEIVAADTLEIPCR 3058
QY 355 -IKW- -D- 358
Db 3059 TEGIPPEISWFLDGKPILEMPGVTVYKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQD 3118
QY 359 -P- -T- -GN- 362
Db 3119 TYVEISEPPRVVMASEWVRVVEGRQTTIRCEVFGNPEPVVNMUKOGEPTYSDLLQSTKL 3178
QY 363 -L- -LA- -SC- -S- -D- -D- -M- 371
Db 3179 SYLHLRETTLADGGTYTCIATNKAGESQTTTDEVLVPPRIEDEERVLOGEGNTYVHC 3238
QY 372 -T- -L- 373
Db 3239 QVTGRPVVVTWKRNGKEIQFNPVHLIRNATRADEGKYSCIASNEAGTAVADFLLDVFT 3298
QY 374 -KI- -W- -S- -M- -K- 379
Db 3299 KPTFETHETTFNIVAGESAKICKIDGHPKPTISWLKGRPFNMMDNIILSPRGDTLMILK 3358
QY 380 -OD- -N- -C- 383
Db 3359 AQRFDGLYTCVATNSYGDSEQDFKVVYTKPYIDETIDOTPKAVAGGEIILKCPVLGNP 3418
QY 384 -V- -H- -DL- -Q- -A- -H-NKE-I 394
Db 3419 TPTVTWKRGDDAVPNDSRHTIVNNYDLKINSVTTEDAGQVSCIAVNEAGNLTHYAAEVI 3478
QY 395 -Y- -TI- -K- -W- -SP- 401
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDSHIT 3538
QY 402 -T- 402
Db 3539 INKAKLSDGGKYTCRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLECPI 3598
QY 403 -G- -P- -T- -NN- 408
Db 3599 SGIPQPDVITKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCIATNRGGKASHD 3658
QY 409 -P- 409
Db 3659 FSLDVLSPPEFDIHGTQFTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLT 3718
QY 410 -N- -A- -N- -L- -ML- -A- -SAS- -F- -D- 421
Db 3719 DNVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKEILSPPVIDISRNDVQPO 3778
QY 422 -S- -T- -VR- -L- -W- 427
Db 3779 VAVNQPTIMRCAVTHGHPFSPISIKWLKNGKEVTDENIRIVEBQGVQLILRTDSHAGKMSC 3838
QY 428 -DV- -D- -R- -G- -I- -C- 434
Db 3839 VAENDAGVKELEMLVDVFTPTPVSVSKSDNFIKALGETITLFCNASGNPYQLKWAKGSL 3898

QY 435 -----I-H-----T-----437
Db 3899 IFDSPGARISLKGARLDIPLHKKTDVGYDTCQALNAAGTSEASVSDVLVPPEINRDGI 3958
QY 438 -----LT-----K-----H-----Q-----442
Db 3959 DMSPLPQAQSLTLQCLAQKQPVQMWTLNGTALTHTSTPGITVADSTFIQINNVSLSLSD 4018
QY 443 -----E-----PV-----Y-----S-----447
Db 4019 KGYYTCAENAVAGSDNLMYNVDDVQAPVINSNGGKQVIEGELAVIECLVEGYAPQVSWL 4078
QY 448 -----V-----A-----449
Db 4079 RGNRVETGVQGVRYVTDGRMLTIEARSLDGIYLCATNEAGSAQAQVTLVLSPKI 4138
QY 450 -----FS-----P-----DG-----454
Db 4139 ITSTPGVLTPSSGKFSPLCAVRGYDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4198
QY 455 -R-Y-----L-----A-----SG-----460
Db 4199 ERHLIYECTAKNDAGADTLTFEPVQTVIAPKISTISGNRYINGSEGTETVKEIESESSEF 4258
QY 461 -----S-F-D-----K-----C-----465
Db 4259 SMSKNGVPLPLSNLIFSEDKLIKILSTRLSQGEYSCTAANKAGNATOKTNLVGVAP 4318
QY 466 -----VH-----I-W-----NT-----471
Db 4319 KIMERPRQVVKHKGQVTLWCEASGVPOPAITWYKDNELLTNTGVDETATTKKSVIFSS 4378
QY 472 -----Q-----TG-----474
Db 4379 ISPSQAGVYTKAENWVASTEEDIDLVMIPPEVPERMNVSTNPRQTVFLSCNATGIPE 4438
QY 475 -----A-----L-----V-----H-----478
Db 4439 PVISWMRDSNIAIQNNEKYQILGTLTAIRNVLPDDGDFYHCIAKSDAGQKIATKRLIVNK 4498
QY 479 -----S-YR-----G-----482
Db 4499 PSRPAPIWECDEKPKKTEYIMDRGDTDPDNPQLLPKQVEDSSLSNGSIAYRCMPGP 4558
QY 483 -----T-----G-----G-----I-----FE-----488
Db 4559 RSRTVLLHAAPOFIVKPKNTTAAIGAVELRCSAAGPPHTTTWAKQGLIEDSKFEIA 4618
QY 489 -----V-C-----490
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAVITCYE 4678
QY 491 -----W-----N-A-----493
Db 4679 RNQAYRGLTWEYNGVPMKPLAGIHFMMNGSLVILDTSSLKEGDELYTCKVRNRRHS 4738
QY 494 -----A-G-----DKV-----G-----AS-----A-----502
Db 4739 IPHLTSAFEGVPEVKTDIKVEVNGSDVLDCEVTSDPLTTHVVWTKNDQKMLDDDAIYV 4798
QY 503 -----S-----D-GSV-----CV-----LD-----L-----R-----K 514
Db 4799 LPNNSLVLLNVEKYDEG-VYKCVASNSIGKAFDDTQNLNVGSSRREAYK 4847

RESULT 13

US-10-369-493-6861
; Sequence 6861, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6861
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-369-493-6861

Query Match 71.2%; Score 2612.1; DB 14; Length 5175;
Best Local Similarity 9.2%; Pred. No. 1.2e-29;
Matches 446; Conservative 53; Mismatches 12; Indels 4339; Gaps 386;
QY 1 M-----SI-----SS-----D-----E-----7
Db 1 MGRSPSNLYGVLLGALLLATTCSVNDKNDPTGKSLAFVFDITGSMFDDLVQVREGAAK 60
QY 8 ---V-----N-----F-L-VV-----13
Db 61 IFKTWQAQREKLIYNYMVPHPDYLGEINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120
QY 14 ---Y-----L-----OE---SG---F-----SHSAF-T 27
Db 121 ILKALQISLPSSFYVFTDARSKDYHLEVLNTIQEKSSVVFVMTGDCGNRTHPGFRT 180
QY 28 -----FG-I---E---S-----H-ISO-----S-NI-----39
Db 181 YEKIAAASFGQVPHLEKSDVSTVLEYRHAHVQKKVHLYEARERGTVSRNTPVDKHL 240
QY 40 ---N-----N-----GA-----LV---P-P---46
Db 241 ELTISLSDKDDSDNLDIVLRDPEGRTVDKRLYSKEGGTIDLNKVLIRLKDPSPGVWTV 300
QY 47 -----A-----A-----LI---S-I-----52
Db 301 NTNSRLKHTIRVFGHGAVDKYGFAASPLDRDLRIELARPRVPLNQDTYLLINMTGLIPGTV 360
QY 53 --I-----Q-----KGL-----Q-Y---VE---AE- 63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPMYFAGPVFPKGLFFVRVQGYDEDNFYEMRIAPT 420
QY 64 ---V-----S-I---N-----ED---GTLF-----73
Db 421 AIGSVIVGGPRAFMSPIHQEFVGRDNLNLSCTVESASAYTIYWKGTEDIIGGFLFYHNTD 480
QY 74 -----D-----G-----R---P-I-----E-----79
Db 481 TSVWTIPELSLKDAGEYECRVISNNGNSVKTRETPESPPEIFGVNRVNSVPLGEAAFLH 540
QY 80 ---S-----L-----SL-I---DA-----86
Db 541 CSTRSAGEVEIRWTRYGATVFNPNTERNTPTNGTLKIHVTRADAGVYECMARNAAGMST 600
QY 87 ---VM-----P-DV-----Q---QAYRD---KLA-----Q-----105
Db 601 RQMRDIMEPPSVKVTTPQDVYFNMRGVNLSCEAMGDPKPEVHWYFKGRLHLLNDYKYQVG 660
QY 93 Q-----T---R---Q---QAYRD---KLA-----Q-----105
Db 661 QDSKFLYIRDTATHDEGTVECRAMSGAQA-RDITDMLATPPKVEIIQNKMMVGRGDRV 719
QY 106 -----Q-H---A---AAA-----AA- 113
Db 720 SPEKTIKPKPKIRWFKNGKDLKIPDDYIKINEQLHMGAKADAGAYSCVGENMAG 779
QY 114 -AAAA-----T---N-----Q-----Q-GS---A-KN- 126

Db 780 KOQVANLSVGRPTTIESPHTVRNIEQVTLQCLAVGIPPEIEWQKGNVLLATLNP 839
QY 127 ---GEN---T--A-N-----132
Db 840 RYTQLADG-NLLJTDQIEDQGFCTIARNTYQQSQSQTLLMVTGLVSPVLGHVPPEEQ 898
QY 133 ---G---EE---N--G---137
Db 899 IEGQDITLSCVVLTGPKPSIVWKDDPKVEEGFTIKIEGGGSLRLRGNGPKDEKYTC 958
QY 138 ---A---H---TI---A--N-NH-----145
Db 959 IAVSPAGNLTLLHINVQLIKKPEFVYKPEGGIVFKPTISGMDKXHAVVNSTHVDLQEGF 1018
QY 146 -----T-----146
Db 1019 AIPCWVSGTPPPIITWYLDGRPTPNSRDFVTADNTLIVRKADKSVSGVYTQATNSAG 1078
QY 147 D---M---E---VD---GD---V-155
Db 1079 DNEQKTTIRIMTPTMISPGQSFMMVVDLFTIPCDVYGDPKFVITWLLDDKPFTEGVN 1138
QY 156 E---I---PS---N--K-A---VVL 165
Db 1139 EDGSLTIPVNEAHRGFTTCHAQNAAGNDRTRVTLVHTTPTTINAENQEKIALQNDIVL 1198
QY 166 -----R-----G--167
Db 1199 ECPAKALPPVRLTWYEGEKIDSQLPHTIREDGALVQNVKLENTGVFCVQVNSLAGED 1258
QY 168 ---HE---SEV---F--I-C-A---WNP-V-180
Db 1259 SLSYTLTVHEKPKIIEVPGVVDWVGFTIEIPCRATGVPEVIRTNKNGIDLKMDEKXF 1318
QY 181 S-D-L-L-V-S--G--SG--D--S--TA---193
Db 1319 SVDNLGLTRIYEADKNDIGNYCVVTNEACTSQWTHVDVQEPPIILPSTQTNNTAVGD 1378
QY 194 RI-----W-----NL---S-ENST---S--204
Db 1379 RVELKCYVEASPASVTFWRFGIAIGTDTKGYVVESDGLVIOASVEDATIYTCRASNP 1438
QY 205 -G-----P--TQ---L--V-----L--211
Db 1439 AGKAEANLQVTVIASPDIKDPDVVTOESIKESHPSFLYCPVFNPLPQISWYLNKPLID 1498
QY 212 ---R-H---CI-R-----216
Db 1499 DKTSWKTDDKRLHVFKAKITDSGVYKCVARNAGEGSKSFQVEVIVPLNLDSEKYYKK 1558
QY 217 ---EG---G---Q---D---V-P-----SN-----225
Db 1559 VFAKEGEVTLGCPVSGFPVQINWVVDGTWVEPGKYKGATLSNDGLTLHLFDSVSVKQE 1618
QY 226 -----KQVT-S-L-L---D---W 233
Db 1619 GNYHCVAQSKNILDIDVELSVLAVPIVGBDDNLEVLFGKDILSLSCDLQTESDDKTTFW 1678
QY 234 -N-SE-----G-----TL--L-----240
Db 1679 SINGESDRPDNVQIPSDGHLRYITDAKPENNGKYMCRVNTNSAGKAERTLTLDVLEPPVF 1738
QY 241 -----A---TG-----SY--D-----246
Db 1739 VEPVEANQKLIKNPILQCOVTGNPKPTVIWKIDGNDVDKSWLFDESLLRIEKLJTG 1798
QY 247 -----G---F---A-----R-----I-251
Db 1799 KSAQISCTAENKAGTASRDFQIONIAAPTFKNEGQETIFRESEITITLDCPVSLGDFQIT 1858
QY 252 W-----T-----KD-----255

Db 1859 WMKQGLPLTENDAIFTLONTRLTILNANRDHEDIYTCVANNTAGVSKDFDVVVQVLPKI 1918
QY 256 -----GN---L--A-----S---260
Db 1919 KNAVTLTLEINEGEEIILTCDAEGNPPTTAKWDFNOGDLPEKFAFVNNHTVVNNVTKYH 1978
QY 261 T---L---GQ---H--K-----266
Db 1979 TGYYKYATNKVGQAVKTINVHRYKPFESGLTESELTVNLTRSLTLECDVDDAIGVGI 2038
QY 267 ---G-P-----I-----F---270
Db 2039 SMTVNGKPLAETDGVQYTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTFNLEVOVP 2098
QY 271 -----AL---K---W-----274
Db 2099 PTIVNEGGYTVIENNSLVLPCVETGKPNPVVTWKDGRPVGDLKSVQVLSQEQQPKIVH 2158
QY 275 ---N---K---KG-----278
Db 2159 ABIAHKGSVICAKNDVGTABISFDVDIITRPMIOKGIKNIVTAIKGGALPFKPCPIDDDK 2218
QY 279 NF-----IL-----S---AG---285
Db 2219 NFKGQIILWRNYQPIDLEAEDARITRLSNDRRLTILNVTENDEGOYSCRVKNDAGENSFD 2278
QY 286 ---V---D--KT--TI---I-W--D-A-H-----296
Db 2279 FKATVLVPFTIIMLDKDNKTAVEHSTVTLSCPATGKPEPDITWFKDGEAHIENIADII 2338
QY 297 ---T---G--EA-----K---Q 302
Db 2339 PNGELNGQLKTRIKEGDAGKYTCEDNSAGSVEQDVNVNVTIIPKIEKDGPSPDESQ 2398
QY 303 Q-----303
Db 2399 QNERVVISCPVYARPPAKITWLKAGKQLSDKFVKTSAWGOKLYLFLKRLRETDSSKYTCIA 2458
QY 304 ---F---P-F-----H---S-----308
Db 2459 TNEAGTKRDKFKVSMVLVAPSFDEPNIVRITVNSGNPSTLHCPAKSPSPFTITWLDGNA 2518
QY 309 ---A-----P-----310
Db 2519 IEBNDYVFFDAGRQIQISKTEGSDQGRYTCIATNSVSGDDLENTLEVIIPPIDGERRE 2578
QY 311 --A---L--D---VD--WQ---S-N-----319
Db 2579 AVAVIEGFSSELPFCDNSTGTVDVWQKDLITINQDTLRGDSFIQIPSSGKKMSFLSARKS 2638
QY 320 -----N---T---FA-----S-----C--S-T---327
Db 2639 DSGRYTCIVRNPAGEARKLDFAVNDPPPSISDELSSANTQITVIPPVPEINCVVSGSPHP 2698
QY 328 ---DM-----C-----I-331
Db 2699 KVTWLFDDKPLEPDSAAVELTNNGETLKIQRVQVEHAGTVTCEAONNVGKARKDFLVRVT 2758
QY 332 ---H---V---C---KL-----G-----337
Db 2759 APPHFEKEREBVAVGDTMLLTCAESSVPLSSVYWHADHESVQNGVITSKYAANEKTL 2818
QY 338 ---O-----DR---PI--K-----343
Db 2819 NVTNIQLDDEGGFYCTAVNEAGITKKFFKLIIVETPYFLDQOKLYPLIILKRLTLDCSAT 2878
QY 344 ---T--F-----345
Db 2879 GTPPPIILFMKCKRLNESDEVDIIGSTLVIDNPQVEGRTYCIENKAGRSEKMDMVE 2938
QY 346 ---Q-----G-H-----348
Db 2939 VLLPPKLSKEWINVEVQAGPLTLECFIEDTSGVHITWRSQFGKQGLDMRAOSSSKSK 2998

QY 349 ---T---N---E---VN---A--- 354
Db 2999 LYIWOATPEDADSYCIAVNDAGAEAVFQVNTTPKIFGDSFSTEIVADTLEIPCR 3058
QY 355 ---IKW--- 358
Db 3059 TEGIPPEISWFLDKGPILFEMPGVYKQGLSLRIDNIKNQBGRYTCAENKAGRAEQD 3118
QY 359 ---P---T---GN--- 362
Db 3119 TYVEISEPRVMASEVMRVEGRQTTIRCEVFGNPEPVNWLKDGEPYTSLLQFSTKL 3178
QY 363 ---L---LA---SC---D---D---D---M--- 371
Db 3179 SYLHLRETTLDGGTCTCIATNAGESQTTTDEVLVPPRIEDEERVLOKEGNTYVHC 3238
QY 372 ---T---L--- 373
Db 3239 QVTGRVPVYVWKRNGKEIQFNPVLHNRATRADEKGYSCIASNEAGTAVADFLIDVFT 3298
QY 374 ---KI---W---S---M---K 379
Db 3299 KPTFETHETFNIVESAKIECKIDGHPKPTISWLKGGPFPNMDNIIILSPRGDTMLK 3358
QY 380 ---QD---N---C--- 383
Db 3359 AQRFDGGLYTCVATNSYGDSEQFKVNYTKPYIDETIDTPKAVAGGEIILKCPVLGNP 3418
QY 384 ---V---H---DL---Q---A---H-NKE-I 394
Db 3419 TPTVTWKRGDVAPENDSRHTVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVI 3478
QY 395 ---Y---TI-K---W---SP--- 401
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYDRYSISPDGSHIT 3538
QY 402 ---T--- 402
Db 3539 INKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYIECPI 3598
QY 403 -G-P---G---T---NN--- 408
Db 3599 SGIPQPDVITKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRVCTCTATNRGGRKASHD 3658
QY 409 ---P--- 409
Db 3659 FSLDVLSPPEFDIHGTOPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKOSRALDGLT 3718
QY 410 -N---A---N---L---ML---A-SAS---F---D--- 421
Db 3719 DNVDISDDGKLTISQASLENAGLYTCIALNRAGEASLEKVELSPPPVIDISRNDVQPO 3778
QY 422 ---S---T---VR---L---W--- 427
Db 3779 VAVNQPTIMRCAVTHGFPFISIKWLKNGKEVTDENIRIVEQGVQLILRTDSHAGKWSC 3838
QY 428 ---DV---D---R-G-I---C--- 434
Db 3839 VAENDAGVKEMLVDVFTPPVSVKSDNPRIKALGETITLFCNAGSNPPQLKWAQKGS 3898
QY 435 ---I-H---T--- 437
Db 3899 IFDSPGARIKGLADIDPHLAKTVDGYTCQALNAAGTSEASVSVLVLPPEINRDGI 3958
QY 438 ---LT---K---H---Q--- 442
Db 3959 DMSPLPAQQLTLQCLAQKGPQMRWTLNGTALTHTSTGTIVASDSTFIQINNVSLSD 4018
QY 443 ---E---PV---Y---S--- 447
Db 4019 KGVYTCYAENVAGSDNLMNVVQAPVISNGGKQVIEGELAVIECLVEGYPAPQVSWL 4078

QY 448 ---Y---A--- 449
Db 4079 RGNRVETGVQGVYVYTDGRMLTIIEARSLDSGILCSATNEAGSAQAQVTLVLSPKI 4138
QY 450 ---PS---P---DG--- 454
Db 4139 ITSTPGVLTPSGSKFSLPCAARGYPDPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4198
QY 455 -R-Y---L---A---SG--- 460
Db 4199 ERHLIYECTAKNDAGADTLEFPVQTVIAPKISTISGNRYINGSEGTETVIKCEIESESEF 4258
QY 461 ---S---P-D---K---C--- 465
Db 4259 SMSKNGVPLPSNNLIFSEDYKLIKILSTRLSQGEYSCTAANKAGNATOKTNLNVGAP 4318
QY 466 ---VH---I-W---NT--- 471
Db 4319 KIMERPRTOVHKGDQVTLWCEASGVPOPAITWYKDNELLTNTGVDDETATTKKSVIFSS 4378
QY 472 ---Q---TG--- 474
Db 4379 ISPSQAGVYTKAENVVASTEEDIDLIVMIPPEVPERMNVSTNPRQTVPLSCNATGIPE 4438
QY 475 ---A---L---V---H--- 478
Db 4439 PVISWMRDSNIAIQNNEKYQILGTLTIAIRNVLPPDDGFYHCIAKSDAGOKIATKLI VNK 4498
QY 479 ---S-YR---G- 482
Db 4499 PSDRPAPIWECEDEKGPKKTEYMDRGDTPDDNPQLPEWKVEDSSLNGSIAYRCMPGP 4558
QY 483 ---T---G---I---FE--- 488
Db 4559 RSRVTLLHAAPQFIVKPKNTAAIGAIVELRCSAAGPPHTTITWAKDGKLIEDSKFEIA 4618
QY 489 ---V-C--- 490
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAPINVDNNILPTPKSSNQKNVAVITCYE 4678
QY 491 ---W---N-A--- 493
Db 4679 RNQAYSRGLTWEYNGVPMKPNLAGIHFMMNGSILVILDTSSLKEGDLLEYTCVKVRNRRHS 4738
QY 494 ---A-G---DKV---G---AS--- 502
Db 4739 IPHLTSAPFGEVPEVKTIDKVENNGSVVLDCVETSDPLTTHVVWTKNQKMLDDDAIYV 4798
QY 503 ---S---D-GSV---CV---LD---L---R---K 514
Db 4799 LPNNSILVLLNVEKYDEG-VYKCVASNSIGKAFDDTQNLNVYGGSSRREAYK 4847

RESULT 14

US-10-120-801-75

; Sequence 75, Application US/10120801

; Publication No. US20030203843A1

; GENERAL INFORMATION:

; APPLICANT: Pena, Carol

; APPLICANT: Guo, Xiaojia

; APPLICANT: Shimkets, Richard

; APPLICANT: Padigar, Muralidhara

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Spytek, Kimberly

; APPLICANT: Mehraban, Fuad

; APPLICANT: Topper, James N.

; APPLICANT: Malyankar, Uriel

; APPLICANT: Wasserman, Scott

; APPLICANT: Edinger, Shlomit

; APPLICANT: Smithson, Glenda

; APPLICANT: Gunther, Erik

; APPLICANT: Komuves, Laszlo

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-340
CURRENT APPLICATION NUMBER: US/10/120, 801
CURRENT FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: 60/285748
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: 60/286068
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/286292
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/288334
PRIOR FILING DATE: 2001-05-03
PRIOR APPLICATION NUMBER: 60/291241
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/322284
PRIOR FILING DATE: 2001-09-14
PRIOR APPLICATION NUMBER: 60/285609
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 75
LENGTH: 5198
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-10-120-801-75

Query Match 71.2%; Score 2610.8; DB 14; Length 5198;
Best Local Similarity 9.2%; Pred. No. 1.3e-29;
Matches 447; Conservative 52; Mismatches 12; Indels 4362; Gaps 387;

QY	1	M	-----SI-----SS-----D-----E-----7
DB	1	MGRSPSWLYGVLGLLLATTCCSSVNDKNDKNDPTKSSLAFVFDITGSMFDDLQVREGAAK 60	
QY	8	V-----N-----F-----L-----VY-----13	
DB	61	IFKTVMAQREKLIYIMVPHDPYLGEIINTTSTYFMRQLSKVYVHGGGDCPEKLTG 120	
QY	14	-----R-----Y-----L-----OE-----SG-----F-----SHSAF-T 27	
DB	121	ILKALQISLPSSFIYFTDARKDYHLEDEVLNTIQEKQSSVVFVMTGCGNRTHPGRT 180	
QY	28	-----FG-I-----E-----S-----H-ISQ-----S-NI-----39	
DB	181	YKIAAASFGVFHLEKSDVSTVLEYVRHAVKQKVHLMYEARERGTVSRNIPVDKHL 240	
QY	40	-----N-----N-----GA-----LV-----P-----P-----46	
DB	241	ELTISLGDKSDNLDIVLRDPEGRVTDKRLYSKEGGTIDLKVKLIRLKDPSGVWTV 300	
QY	47	-----A-----A-----A-----LI-----S-----I-----52	
DB	301	NTNSRLKHTIRVPFGHNAVDFKYGFASRPDLRIELARPRVPLNODTYLLINMTGLIPGTV 360	
QY	53	--I-----Q-----KGL-----Q-Y-----VE-----AE-63	
DB	361	GEIDLVDYHGHSLYKAVASPHRTNENMYFAGFPVPGKGLFFVRVQGYDEDNVEFMRIAPT 420	
QY	64	--V-----S-I-----N-----ED-----GTLF-----73	
DB	421	AIGSVIVGGPRAFMSPHQBFVGRDLNLSCTVESASAYTIYVWKTGEDIIGGFLVHNTD 480	
QY	74	-----D-----G-----R-----P-----I-----E-----79	
DB	481	TSVMTPELSKDAGEVECEVSNNGNYSVKTRVETRESPEIFGVRNVSVPLGEAFLH 540	
QY	80	-----S-----L-----I-----DA-----86	
DB	541	CSTRSAGEVEIRTRYGATVFNENPTNPTNGTLKLHHVTRADAGVYECMARNAGMST 600	
QY	87	-----VM-----P-DV-----V-----92	
DB	601	RKMRLDIMEPPSVKVTQDVIYFNNREGVNLSCAMGDKPEVHWYFKGRLLNDYKYQVG 660	

QY	93	Q	-----T-----R-----Q-----QAYRD-----KLA-----Q-----105
DB	661	QDSKFLYIRDATHHDEGTVECRAMSQAGQA-RDITDLMLATPPKVEIIQNKMVGRGDRV 719	
QY	106	-----Q-H-----A-----AAA-----AA-113	
DB	720	SPECKTIRGPHPKIRWFKNGKDLIKPDDYIKINEQOLHMGAKODAGAYSCVGENMAG 779	
QY	114	--AAAA-----T-----N-----Q-----Q-GS-----A-KN-126	
DB	780	KDVQVANLSVGRVPTIESPHTRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNPP 839	
QY	127	-----GEN-----T-----A-N-----132	
DB	840	RYTQLADG-NLLITDAQIEDQOFTCIARNITYQQSQSQTLLMVTGLVSPVLGHVPPEQL 898	
QY	133	-----G-----E-----N-----G-----137	
DB	899	IEGQDLTSCVVVLGTPKPSIVMIKDDKPVREBPTIKIEGGGSLLRGNPKDEKGYTC 958	
QY	138	-----A-----H-----TI-----A-N-NH-----145	
DB	959	IAVSPAGNSTLHINVOLIKKPEFVYKPEGIVFKPTISGMDKXHAVVNSTHVDLGEGF 1018	
QY	146	-----T-----146	
DB	1019	AIPCVCVSGTPPPIITWYLDGRPIITNSRDFTVTADNTLIVRKADKSYSGVYTCQATNSAG 1078	
QY	147	D-----M-----M-----E-----VD-----GD-----V-155	
DB	1079	DNEQKTTIRIMTPIISPGQSSFNVMVDDLFTIPCDVYGDPKPVITWLDDDKPFTGTVN 1138	
QY	156	E-----I-----PS-----N-K-A-----VVL 165	
DB	1139	EGSLTIPVNEAHRGFTTCHAOAAGNDTRVTLVHTTPTINAENQEKIALQNDIVL 1198	
QY	166	-----R-----G-----167	
DB	1199	ECPAKALPPPVLWTVEGEKIDSLIPHTIREDGALVQNVKLENTGVFVQCVSNLAGED 1258	
QY	168	-----HE-----SEV-----F-----I-C-A-----WNP-V-----180	
DB	1259	SLSYTLTVHEKPKIIESEVPGVVDWVKGFTEIPCRATGVPEVIRTNKNGIDLKMEKKF 1318	
QY	181	S-D-L-L-V-S-----G-----SG-----D-----S-----TA-----193	
DB	1319	SVDNLTGLRIYEADKNDIGNYCVTVNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVD 1378	
QY	194	RI-----W-----NL-----S-ENST-----S-204	
DB	1379	RVELKCVVEASPPASVTFRRGIAIGTDTKGVVVEDSGTLVIQASVEDATIVTCKASNP 1438	
QY	205	-G-----P-----TQ-----L-----V-----L-----211	
DB	1439	AGKAEANLQVTVIASPDIKDPDVVTQESIKESHPSLYCEVFSNPLFOISWYLNKPLID 1498	
QY	212	-----R-H-----CI-R-----216	
DB	1499	DKTSWKTSDDKRLHVFKAKITDSGVYKCVARNAAGESKSFQEVIVPLNLDESXYKKK 1558	
QY	217	-----EG-----G-----Q-----D-V-P-----SN-----225	
DB	1559	VPAKEGEVTLGCPVSGFPVPQINWVVDGTVVBPBKYGKATLSNGLTLHPDSVSVKQE 1618	
QY	226	-----KQVT-S-L-----D-----W-233	
DB	1619	GNVHCVAQSKGNILIDIVELSVLAVPIVGEDDNLVFLGKDISLSCDLQTESDKTTFW 1678	
QY	234	--N-SB-----G-----TL-----L-----240	
DB	1679	SINGSSEDRPDNVQIIPSDGHRLYITDAKPNNGKYMCRVTNSAGKAERTLTLDVLEPPVF 1738	
QY	241	-----A-----TG-----SY--D-----246	

Db 1739 VEPVFANQKLGNNPIILOQVGTGNPKPTVIWKIDGNDVDKSWLFDESLSLRIBKLTG 1798
QY 247 -G- -F- -A- -R- -I- -251
Db 1799 KSAQISCTAENKAGTASRDFQIQAAPTAKNEGQSTIFRESEITLDCPVSLGFQIT 1858
QY 252 W- -T- -KO- -255
Db 1859 WMKQGLPLTENDAITFLDNTRLTILNANRDHDIYTCVANNTAGQVSKDFVVQVLPKI 1918
QY 256 -GN- -L- -A- -S- -260
Db 1919 KNAVVTLEINEGEIILTCDAENPTPTAKWDFNQGDLPKEAFVANNHTVVVVNNVTKYH 1978
QY 261 T- -L- -GO- -H- -K- -266
Db 1979 TGVIKCVATNKVQAVKTVINHVRTKPRFESGLTESELTVNLTRSLTECDVDDAIGVGI 2038
QY 267 -G- -P- -I- -F- -270
Db 2039 SWTVNGKPFLEAETDGVQTLAGRFLHIVSAKTDHGSYACTVTNEAGVATKTFNLFQVP 2098
QY 271 -AL- -K- -W- -274
Db 2099 PTIVNEGGEYTVIENNSLVLPCVETGKPNPVVWTGDRPVGLKSVQVLSGQQFKIVH 2158
QY 275 -N- -K- -KG- -278
Db 2159 AEIAHGSYICMAKNDVGTABISFVDIITRPMIQKINIVTAIRGGALPFCRPIDDDK 2218
QY 279 NF- -IL- -S- -AG- -285
Db 2219 NFKGQIILWRYQPIDLEADARITLSNDRRLTILNVTENDEGQYSCRVDNAGENSFD 2278
QY 286 -V- -D- -KT- -TI- -I- -W- -D- -A- -H- -296
Db 2279 FKATVLVPPTIIMLDKDKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAIIHENTADII 2338
QY 297 -T- -G- -EA- -K- -Q- -302
Db 2339 PNGELNGNOLKITRIKEGDAGKYTCADNSAGSVEQDVNVNVTIPKIEKDGPSPYESQ 2398
QY 303 Q- -F- -304
Db 2399 QNERVISCVPYARPPAKITWLKAGPLQSKFVKTSANGQKLYLPLRETDSSKTYCIA 2458
QY 305 -P- -F- -H- -S- -308
Db 2459 TNEAGTDKDRDFKVSMLVAFSDFEPNIVRRITVNSGNPSTLHCPAKGSPSTITWLKDUNA 2518
QY 309 -A- -P- -310
Db 2519 IEPNDRYVFDAGRLQISKTEGDOGRYTCIATNSVSGDDLENTLEVIIPVIDGERRE 2578
QY 311 -A- -L- -D- -VD- -WO- -S- -N- -319
Db 2579 AVAVIEGFSSELCDSNSTGVVEWQKDLTINQDLRGDSFIQIPSSGKKMSFLSARKS 2638
QY 320 -N- -T- -FA- -S- -C- -S- -T- -327
Db 2639 DSGRYTCIVRNPAEARKLFDFAVNDPPSISDELSSANTQITVIPPVVEINCVSQSPHP 2698
QY 328 -DM- -C- -I- -331
Db 2699 KVTWLFDDKPLEPDSAAYELTNGETLKIVRQVEHAGTVTCBAQNNVGKARKDFLVRVT 2758
QY 332 -H- -V- -C- -KL- -G- -337
Db 2759 APHFKEREEVVARVGDWMLTLCNAESSVPLSSVYVHHADESQVQGVITSKYAANEKTL 2818
QY 338 -Q- -DR- -PI- -K- -343

Db 2819 NVTNIQLDDGEFYCTAVNEAGITKFKFLIVIETPYFLDQKQYPIILGKRULTDCSAT 2878
QY 344 -T- -F- -345
Db 2879 GTPPPPIILFWKDGKRLNESDEVIIIGSTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVE 2938
QY 346 -Q- -G- -H- -348
Db 2939 VLLPPKLSKEWINVEVOAGDPLTECPIDETSGVHITWSRQFGKQDGLDMRAQSSSDSKS 2998
QY 349 -T- -N- -E- -VN- -A- -354
Db 2999 LYIMOATPEDADSYSCIAVDAGBAEFQVTVNTPPKIFGDSFSTTEIVADTTLPIPCR 3058
QY 355 -IKW- -D- -358
Db 3059 TEGIPPELSWFLDGKPILEMPGVTYKQGDLSLRIDNIKNQEGRYTCVAENKAGRAEQD 3118
QY 359 -P- -T- -GN- -362
Db 3119 TYVEISEPFRVVMASEVMRWVEGRQTTIRCEVFGNPEPVVWMLKDGEPYTSDDLQFSTKL 3178
QY 363 -L- -LA- -SC- -S- -D- -D- -M- -371
Db 3179 SYLHLRETTLADGGTYTCIATNKAESQTTTDDVEVLVPPRIEDEERVLOGKEGNTYMHVC 3238
QY 372 -T- -L- -373
Db 3239 QVTGRPVVYVTKRNGKEIEQPNVLHNRATRADEGKYSCIASNAGTAVADFLIDVFT 3298
QY 374 -KI- -W- -S- -M- -K- -379
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGGPFNMNDIILSPRGDTLMILK 3358
QY 380 -QD- -N- -C- -383
Db 3359 AQRFQDGLYTCVATNSYSGDEQDFKNVVTYKPIDETIDQTPKAVAGGEIILKCPVLGNP 3418
QY 384 -V- -H- -DL- -O- -A- -H- -NKE- -I- -394
Db 3419 TPTVTWKRGGDVAVPNDNRHTIYNNYDLKINSVTEDAGQYSCIAVNEAGNLTHYAAEVI 3478
QY 395 -Y- -TI- -K- -W- -SP- -401
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYLDYRISISPDGSHIT 3538
QY 402 -T- -402
Db 3539 INKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPI 3598
QY 403 -G- -P- -T- -NN- -408
Db 3599 SGIPQPDVITWTKNGMDINMTDSRVILAQNNTFGIENVQVTDQGRYTCATNRGKASHD 3658
QY 409 -P- -409
Db 3659 FSLDLVSPPEFDIHGTQPTIKREGDTITLTCPKLAEDTADQVMDVSVWTKDSRALDGLT 3718
QY 410 -N- -A- -N- -L- -ML- -A- -SAS- -F- -D- -421
Db 3719 DNVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPVVIDISRNDVQPO 3778
QY 422 -S- -T- -VR- -L- -W- -427
Db 3779 VAVNOPTIMRCVATGHPFPSIKWLKNGKEVTDDENIRIVEQGVQLQILRTSDSHAGKWSC 3838
QY 428 -DV- -D- -R- -G- -I- -C- -434
Db 3839 VAENDAGVKELEMLVDVFTPPVSVKSDNPRIKALGETITLFCNAGSNPYPQLKWKAGGSL 3898
QY 435 -I- -H- -T- -437
Db 3899 IFDSPGARISLKGARLDIPHLKKTVDGVTCTQALNAGTSEASVSVDVLVPPPEINRDI 3958

QY 438 -----L-T-----K-----H-----Q----- 442
Db 3959 DMSPLPAQSLTLQCLAQKPPQMRWTLNGTALTHTSPGTVASDSFIQINNVLSLD 4018
QY 443 -----E-----PV-----Y-----S----- 447
Db 4019 KGVTCTAENVAGSDNLNMYVDVQAPVISNGGTQKVIEGELAVIECLVEGYPAPQVSWL 4078
QY 448 -----V-----A----- 449
Db 4079 RGNRVETGVGVRYTGDGRMLTIIEARSLDSGIYLCSATNEAGSAQAQYTLVLSPKI 4138
QY 450 -----FS-----P-----DG----- 454
Db 4139 ITSTPGVLPSSGSKFSLPCAARGVDPPIISWTLNGNDIKDGENGTIGADGTLHIEKAE 4198
QY 455 -R--Y-----L-----A-----SG----- 460
Db 4199 ERHLIVECTAKNADAGDTLEFPVQTTIVAPKISTSGNRYINGSEGTETVIKCEIESESEF 4258
QY 461 -----S-----F--D-----K-----C----- 465
Db 4259 SMSKNGVPLPSNLLIFSEDYKLIKILSTRLSDQGEYSCTAANKAGNATQKTNLVGVAP 4318
QY 466 -----VH-----I--W-----NT----- 471
Db 4319 KIMERPTQVHKGDQVTLWCEASGVPOPAITWYKDNELLTNGVDETATTKKSVIFSS 4378
QY 472 -----TG-----Q----- 474
Db 4379 IPSQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMVNSTPQTVPFLSCNATGIPE 4438
QY 475 -----A-----L-----V-----H----- 478
Db 4439 PVISWMRDSNIAIQNNEKYQILGTLTALRNVLPPDDGFFHYCIAKSDAGQKIATRKILVAK 4498
QY 479 -----S--YR--G-- 482
Db 4499 PSRPAPIWVECEDEKGPKEKTEYMWIDRGDTPDDNPQLLPHKDVEDSSLNGSIAYRCMPGP 4558
QY 483 -----T-----G-----G-----I-----FE-- 488
Db 4559 RSSRTVLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPPHPTITWAKDGKLIBDSKFEIA 4618
QY 489 -----V--C-- 490
Db 4619 YSHLKVTLNSTSDSEYTCMAQNSVGSSTVSFAFVNDNNILPTPKPSSNOKNVAITCYE 4678
QY 491 -----W-----N--A----- 493
Db 4679 RNQAYSRLTWEYNGVMPKXNLAGIHPMNGSLVILDTSSLKEGDLELYTCKVRNRRHS 4738
QY 494 -----A--G-----DKV-----G-----AS----- 502
Db 4739 IPHLTSAFEGVPEKTIKDVENVGSDVLDCEVTSDELTHVVTWKDQKMLDDDAIYV 4798
QY 503 -----S-----D--GSV--CV-----L-----D-----L----- 512
Db 4799 LPNNSVLLNVEKYDEG-VYKCVASNSIGKAFDDTQLNVYEGDFLPLTGFEGSGINIDDS 4857
QY 513 -----R--K 514
Db 4858 SNAGSSRRREAYK 4870

RESULT 15
US-10-369-493-6858
; Sequence 6858, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6858
; LENGTH: 5198
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-369-493-6858

Query Match 71.2%; Score 2610.8; DB 14; Length 5198;
Best Local Similarity 9.2%; Pred. No. 1.3e-29;
Matches 447; Conservative 52; Mismatches 12; Indels 4362; Gaps 387;
QY 1 M-----SI-----SS-----D-----E----- 7
Db 1 MGRSPSLVGLVGLLLATTCSVNDKNDPTGKSLAFVFDITGSMFDDLVQREGAAK 60
QY 8 ---V-----N-----F--L--VY----- 13
Db 61 IPKTWMAQREKLIYNYVPHDPYLGEIINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120
QY 14 ---Y-----R--Y-----L--QE--SG--F-----SHSAP-T 27
Db 121 ILKALQISLPSSFIYVFTDARSKDVHLEDEVLTIQEKQSSVVFVMTGDCGNRTHPGFRT 180
QY 28 -----PG-I--E--S-----H--SQ-----S-NI----- 39
Db 181 YEKIAAASFGQVHLEKSDVSTVLEYVRHAVKQKVHLYMARERGTVSRNIPVDKHL 240
QY 40 ---N-----N-----GA-----LV--P--P----- 46
Db 241 ELTISLGDKDDNDLIVLRDPEGRTVDKRLYKSGGTIDLKNVKILRLKDPSPGVWTV 300
QY 47 -----A-----A-----LI--S--I----- 52
Db 301 NTNSRLKHTIRVFGHGAVDKYGAFASRLDRIELARPRVLNQDTYLLINMTGLIPPGTV 360
QY 53 --I-----Q-----KGL--Q--Y--VB--AE-- 63
Db 361 GEIDLVDYHSHLYKAVASPHRTNPNMYFAGPVPKGLFFVVRVQGYDEDNYEFMRIA 420
QY 64 ---V-----S--I-----N-----ED--GTLF----- 73
Db 421 AIGSVIVGGPRAFPMSPIHQEFVGRDLNLSCTVESASAYTIYWKGTEDIIGGFLFYHNTD 480
QY 74 -----D-----G-----R--P--I-----E----- 79
Db 481 TSVMTIPELSLKDAGEYECRVISNNNGYSVKTRVETRESPPPEIFGVVRNVSVPLGEAFLH 540
QY 80 -----S-----L-----SL--I-----DA----- 86
Db 541 CSTRSAGEVEIRWTRGVATVNGPNTERNPTNGTLKIHVTRADAGVYECMARNAGMST 600
QY 87 -----VM-----P--DV-----V-- 92
Db 601 RKMRLDIMEPPSVKVTTPQDVFYFNMREGVNLSCAMGDKPKVEVHWYFKGRHLLNDKYQVG 660
QY 93 Q-----T--R--Q--QAYRD-----KLA-----Q----- 105
Db 661 QDSKFLYIRDATHHDEGTVECRAMSQAQA-RDTTDLMLATPPKVEIIONKMMVGRGDRV 719
QY 106 -----Q--H--A-----AAA-----AA-- 113
Db 720 SFEKTIKPKPKIRWFKNGKDLIKPDDYIKINEGQLHIMGAKDEBAGAYSCUGENMAG 779

Db 2939 VLLPPLSKWENVEQADPLTLECPDTSVGHITWSRQFGKQGLDMRAQSSDKSK 2998
QY 349 -T- -N- -E- -VN- -A- - 354
Db 2999 LYINQATPEDADSYCIAVNDAGAEAVQVNTPPKIFGDSFSTTEIVADTLEIPCR 3058
QY 355 -KW- -P- -S- -D- - 358
Db 3059 TEGIPPEISWFLDGKPILEMPGYTKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQD 3118
QY 359 -P- -T- -GN- - 362
Db 3119 TYEISBPVRVMASEVMRVEGQRTIRCEVFCNPPVNWLLKDGEPYTSLLQPSKTL 3178
QY 363 -L- -LA- -SC- -S- -D- -D- -M- - 371
Db 3179 SYLHRETTLAGGTYTCIATNKAGESQTTDVEVLVPPRIEDBEERVLOGKEGNTVMVC 3238
QY 372 - 373
Db 3239 QVTCRPVYVTKENGKEIQFNVLHNRATRADEKYSCIASNEAGTAVADFLIDVFT 3298
QY 374 -KI- -W- -S- -M- -K 379
Db 3299 KPTFETHETFNIVEGESAKIECKIDGHPKPTISWLKGRPFENNIIILSPRGDTLMILK 3358
QY 380 -OD- -N- -C- - 383
Db 3359 AQRFDGLYTCVATNSYGDSEQDFKNVYTKPYIDETIDQTPKAVAGEIILKCPVLGNP 3418
QY 384 -V- -H- -DL- -Q- -A- -H-NKE-I 394
Db 3419 TPTVTWKRGDVAPENDSRHTIVNNDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVI 3478
QY 395 -Y- -TI-K- -W- -SP- - 401
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYLDYRYSISPDGSHIT 3538
QY 402 -T- - 402
Db 3539 INKALSDGGKYICRASNEAGTSDIDILKILVPPKIDKNIIGNPLAIVARTIYLECPI 3598
QY 403 -G- -P- -T- -NN- - 408
Db 3599 SGIQPDVITWKGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCATNREGKASHD 3658
QY 409 -P- - 409
Db 3659 FSLDVLSPFPDIHGTOPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKDSRALDGLT 3718
QY 410 -N- -A- -N- -L- -ML- -A- -SAS- -F- -D- - 421
Db 3719 DNVDISDDGRKLITISQASLENAGLYTCIALNRAGEASLEPKVILSPFVIDISRDVQPO 3778
QY 422 -S- -T- -VR- -L- -W- - 427
Db 3779 VAVNQPTIMECAVTHGHPFPIKWLKNGKEVTDENIRIVEQGVQLILRTDSHDAGKWSC 3838
QY 428 -DV- -D- -R- -G- -I- -C- - 434
Db 3839 VAENDAGVKEMLVDVFTPPVSVKSDNPICALGETITILFCNASGNPYPLKWKAGGSL 3898
QY 435 -I- -H- -T- - 437
Db 3899 IFDSPGARIKLGARLDIIPHILKKTVDGYTCQALNAGTSEASVSDVLVPPEINRDGI 3958
QY 438 -LT- -K- -H- -Q- - 442
Db 3959 DMSPLPAQOQSLTLQCLAQCKPVPQMRWTLNGTALTHTSTPGITVASDSTFIQINNVSLS 4018
QY 443 -E- -PV- -Y- -S- - 447
Db 4019 KGVYTCVAENVAGSDNLMYNDVVVQAEVINSNGGKQVIEGELAVIECLVEGYPAPQVSWL 4078

QY 448 -V- -A- - 449
Db 4079 RGNRVETGVQVRYVTDGRMLTIIIEARSLDSGIYLCSATNEAGSAQAYTLEVLVSPKI 4138
QY 450 -FS- -P- -DG- - 454
Db 4139 ITSTPGVLTPSSGSKESLPCAVRGYPDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4198
QY 455 -R- -Y- -L- -A- -SG- - 460
Db 4199 ERHLIYEYAKNDAGADTLEFPVQTIIVAPKISTSGNRYNGSEGTETVIKCEIESSESEF 4258
QY 461 -S- -F- -D- -K- -C- - 465
Db 4259 SMSKNGVPLLPNNLIPFSEDIKLIKILSTRLSQGBYSCTAANKAGNATQKTNLVGVAP 4318
QY 466 -VH- -I- -W- -NT- - 471
Db 4319 KIMERPTQVVKGDQVTLWCEASGVPOPAITWYKDNELLTNTGVDETATTTKKKSIVFSS 4378
QY 472 - 474
Db 4379 ISPSQAGVYTCKAENWVASTEEDILVIMIPPEVWPERMNVSTNPRQTVFLSCNATGIPE 4438
QY 475 -A- -L- -V- -H- - 478
Db 4439 PVISWRDSNIAIQNNEKYQILGTLTILAIRNVLPDDGDFYHCIAKSADAGOKIATRKLIUNK 4498
QY 479 - 482
Db 4499 PSRPAPIWECDEKPKKTEYMDRGDTPDDNPOLLPWKOVEDSSLANGSIAYRCMPGP 4558
QY 483 -T- -G- -I- -FE- - 488
Db 4559 RSSRTVLLHAAPOQIVPKNTTAAIGAIVELRCSAAGPPHPTITWAKDGKLIEDSKFEIA 4618
QY 489 - 490
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAPINVDNNILPTPKPSSNQKNVAVITCYE 4678
QY 491 -W- -N- -A- - 493
Db 4679 RNQVSRGLTWYGVMPKPNLAGIHFMMNGSLVILDTSSLKBDLELVTCVKVNRNRHS 4738
QY 494 -A- -G- -DKV- -G- -AS- -A- - 502
Db 4739 IPHLTSAFEGVPEVKTIIDKVEVNNGDSVVLDCEVTSPLTTHVVTKNQKMLDDAIYV 4798
QY 503 -S- -D- -GSV- -CV- -L- -D- -L- - 512
Db 4799 LPNNSLVLLNVEKYDEG-VYKCVASNSICKAFDDTQNLNVYEGDFLPTGTGEGSGINIDDS 4857
QY 513 -R- -K 514
Db 4858 SNAGSSRREAYK 4870

Search completed: January 3, 2005, 15:49:10
Job time : 139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:09:41 ; Search time 21.6667 Seconds
(without alignments)
2282.558 Million cell updates/sec

Title: US-09-987-701-12
Perfect score: 3669
Sequence: 1 MSISSDEVNFLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM30
Gapop 1.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: Pir1: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2612.1	71.2	5175	2 T20992	hypothetical prote
2	2610.8	71.2	5198	2 T43290	hemocentin precurs
3	2603.7	71.0	7962	2 I38346	elastic titin - hu
4	2594.2	70.7	6658	2 T13931	projectin - fruit
5	2593.7	70.7	10797	2 T30192	probable peptide s
6	2592.2	70.7	26926	1 I38344	titin, cardiac mus
7	2590.9	70.6	7829	2 T15789	hypothetical prote
8	2587.7	70.5	4545	1 S25111	alpha-2-macroglobu
9	2586.1	70.5	4544	1 S02392	alpha-2-macroglobu
10	2585	70.5	6831	2 A88852	protein unc-22 [im
11	2585	70.5	6839	2 S57242	twitchin [similar]
12	2585	70.5	7160	2 T27935	hypothetical prote
13	2583.7	70.4	4660	2 T42377	gp330 protein prec
14	2580.1	70.3	5825	2 T12117	polyprotein - fava
15	2579.8	70.3	3461	2 S58870	reelin precursor -
16	2579.7	70.3	5376	2 T42215	zonadhesin - mouse
17	2575.1	70.2	8243	2 T31307	type I fatty acid
18	2570.7	70.1	4543	1 A53102	alpha-2-macroglobu
19	2569.3	70.0	5232	2 A45086	HC-toxin synthetas
20	2566	69.9	6805	2 S20901	titin - rabbit (fr
21	2556.6	69.7	4753	1 A47437	LDL-receptor-relat
22	2550.8	69.5	15281	2 S41309	cyclosporin synthe
23	2548.4	69.5	4302	2 A38971	polycystic kidney:
24	2547.9	69.4	6669	2 S55024	nebulin, skeletal
25	2545	69.4	4767	2 T31345	hypothetical prote
26	2543	69.3	6642	2 T29757	protein UNC-89 - C
27	2542.1	69.3	5138	2 B96695	hypothetical prote
28	2541.3	69.3	4861	2 S71752	giant protein p619
29	2538.8	69.2	3623	2 T09456	intrinsic factor-B

30 2538.3 69.2 9376 2 T14593 syringomycin synth
31 2536.9 69.1 4836 2 T14346 herc2 protein - mo
32 2528.1 68.9 4930 2 E69679 polyketide synthet
33 2526.8 68.9 4447 2 A69679 polyketide synthas
34 2525.5 68.8 3856 2 T51174 ataxia-telangiecta
35 2523.8 68.8 5369 2 T44807 mycosubtilin synth
36 2523.5 68.8 4644 1 A38905 dynein heavy chain
37 2523.2 68.8 4572 2 S57908 hypothetical 527K
38 2522.6 68.8 4568 2 T08030 dynein beta heavy
39 2520.2 68.7 4639 2 A54794 dynein heavy chain
40 2519.9 68.7 3623 2 T08618 intrinsic factor-B
41 2519 68.7 4344 1 A53489 dynein heavy chain
42 2513.1 68.5 4464 2 D87755 protein T21E12.4 f
43 2511.7 68.5 4391 2 A38096 perlecan precursor
44 2511.1 68.4 6486 2 T31076 tyrocidine synthet
45 2508.3 68.4 4273 2 C69679 polyketide synthas

ALIGNMENTS

RESULT 1

T20992 hypothetical protein F15G9.4a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20992; T24733

R:Sulston, J.
submitted to the EMBL Data Library, December 1994

A:Reference number: Z19355
A:Accession: T20992

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5175 <WIL>

A:Cross-references: EMBL:Q810L3; EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a

A:Experimental source: clone F15G9

R:Kershaw, J.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z19929

A:Accession: T24733

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5175 <W12>

A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a

A:Experimental source: clone T09B9

C:Genetics:

A:Gene: CESP:F15G9.4a

A:Map position: X

A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
; 2512/2; 2591/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/1

Query Match 71.2%; Score 2612.1; DB 2; Length 5175;

Best Local Similarity 9.2%; Pred. No. 6e-53;

Matches 446; Conservative 53; Mismatches 12; Indels 4339; Gaps 386;

Qy	1	M-----SI-----SS-----D-----E-----	7
Db	1	MGRSPSWLVGLGLLLLTTCSSVNDKNDPTGKSLAFVFDITGSMFDDLQVREGAAK	60
Qy	8	----V-----N-----F-----L-----VY-----	13
Db	61	IPKTVAQREKLIYIMVFFHPDPLGEIINTDTSYFMRLSKVYVHGSGDCEPKLTG	120
Qy	14	-----R-----Y-----L-----OE-----SG-----F-----SHSAF-T	27
Db	121	ILKALQISLPSSFIYVFTDARSKDYLHLEVLNTIQEKSSVVFVMTGDCGNRTHPGFRT	180
Qy	28	-----EG-I-----E-----S-----H-ISO-----S-NI-----	39
Db	181	YEKIAAASFGQVPHLEKSDVSTVLEYVRHAVKOKVHLMYAEARGGTVSRNIPVDKHL	240
Qy	40	-----N-----GA-----LV-----P-----P-----	46

Db	241	ELTISLGDKDDSDNLDIVLRDEGRVTKRLYSKEGGTIDLRKVKLRLKDPGWTW	1300
Qy	47	-----A-----A-----LI--S--I-----	52
Db	301	NTNSRLKHTIRVFGHGAVDKPKYFASRPLDRIELARPRVLNQDTYLLINMTGLIPPGTV	360
Qy	53	-----I-----Q-----KGL-----Q-Y-----VE-----AE-63	
Db	361	GEIDLVDYHSHLYKAVASPHRTNPNMYFAGPEVPKGLFFVRVOGYDNEYFMRAP	420
Qy	64	-----V-----S-I-----N-----ED-----GTLF-----73	
Db	421	AIGSVIUGBPRAFMSPILHQBFVGRDLNLSCTVESASAYTIYWKTEGDIIGGLFVHNTD	480
Qy	74	-----D-----G-----R-----P-I-----E-----79	
Db	481	TSVMTPELSLKDAGEYECRVSNNNGYVSKTRVETRESPEIFGVRNVSVPLGEAAFLH	540
Qy	80	-----S-----L-----SL-I-----DA-----86	
Db	541	CSTRSAGEVEIRWTRYGATVNGENTERNPTNGTLKIHVTRADAGVYECMARNAGMST	600
Qy	87	-----VM-----P-DV-----QAYRD-----Q-----105	
Db	601	RKMRDLIMEPPSVKVTPODYVFNMRGCVNLSCEAMGDPKEVHMYFKGRHLLNDYKYQVG	660
Qy	93	-----T-----R-----Q-----KLA-----105	
Db	661	QDSKFLYIRDATHHDEGTECRAMSOQA-RDITDLMLATPPKVEIIQNMVGRDRV	719
Qy	106	-----S-----L-----Q-H-----A-----AAA-----113	
Db	720	SFECKTIRGKPKIRFKNGKOLIKPDYKIKNEGQLHMGAKDEDAAYSCVGENMAG	779
Qy	114	-----AAAA-----T-----N-----Q-----Q-GS-----A-KN-126	
Db	780	KDVQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNNP	839
Qy	127	-----GEN-----T-----A-N-----132	
Db	840	RYTQADG-NLLITDAQIEDQGOFTCIARTNYGQSQSOTTLMTVLGVSPLVGHVPPBEOL	898
Qy	133	-----G-----G-----EE-----N-----G-----137	
Db	899	IEGQDLTLCVVVLGTPKPSIIVIKDDKPVEEGPTIKIEGGSLRLRGNGPKDEGYTC	958
Qy	138	-----A-----H-----TI-----A-N-NH-----145	
Db	959	IAVSPAGNSTLHINVQLIKKPEFVYKPEGGIVFKPTISGMDEKHVAVVNSTHVDLDEGF	1018
Qy	146	-----T-----146	
Db	1019	AIPCWVGTPPIITWYLDGRPIPTNSRDFVTADNTLIVRKADKSYSGVYTCOATNSAG	1078
Qy	147	D-----M-----M-----E-----VD-----GO-----V-155	
Db	1079	DNEQKTIIRIMNTWISFGQSSFNWVDDLFTIPCDVYGPKPVTIWLDDKPPTEGVN	1138
Qy	156	E-----I-----PS-----N-----K-A-----VVL-165	
Db	1139	EDGSLTIPNVNEAHRGTFTCHAQNAAGNDTRVTLTVHTTPTINAENQEKIALQNDIVL	1198
Qy	166	-----R-----167	
Db	1199	ECPAKALPPPVRMLTYEKEKIDSQLIPIHTIREDGALVLQNVKLENTGVFCQVSNLAGE	1258
Qy	168	-----HE-----SEV-----F-----I-C-A-----WNP--V-----180	
Db	1259	SLSYTLTVHEKPKIIESEVPGVVDVVKGTIEIPCRATGVPEVIRTNKNGIDLKWDEKKE	1318
Qy	181	S-D-L-L-V-S-----G-----SG-----D-----S-----TA-----193	
Db	1319	SYDNLGTLRIYEADKNDIGNYNCVVTNEAGTSQMTTHVDVQBPPIILPSTQNTTAVGSD	1378
Qy	194	RI-----W-----NL-----S-ENST-----S--204	
Db	1379	RVELKCYVEASPPASVTFRRGIAIGTDTKGYVVEDSGTLVIOASVEDATITCKASNP	1438
Qy	205	-G-----P-----TQ-----L-----V-----L--211	
Db	1439	AKAEANLOVTVIASPDIKDPDVVTVQESIKESHPSLYCPVFSNPLPQISWYLNKPLD	1498
Qy	212	-----R-----H-----CI-R-----216	
Db	1499	DKTSWKTSDKRLHVKAKITDGSVYKCVARNAAEGSKSPQVEVIVPLNLDSEKYYK	1558
Qy	217	-----EG-----G-----Q-----D-----V-P-----SN-----225	
Db	1559	VFAKEGEVTLGCPVSGFPVPQINNVGTVVPEPGKYKAGATLSNDGLTLHFDVSUVKQE	1618
Qy	226	-----KDVTS--S--L-----D-----W-233	
Db	1619	GNVHCVAQSKGNLIDIVELSVLAVPIVGEDDNLVFLGKDISLSCDLQTESDDKTTFVW	1678
Qy	234	-N-SE-----G-----G-----TL--L-----240	
Db	1679	SINGSEDRPDNVQIPSDGHRLYITDAKPENNGKYMCRVNSAGKAERTLTLDVLEPPVF	1738
Qy	241	-----A-----TG-----SY--D-----246	
Db	1739	VEPVFEANOKLIGNNPIILOQVTPGPKPTVIWKDNDVDKSWLFDSELSLLRIEKLIG	1798
Qy	247	-----G-----F-----A-----R-----I-251	
Db	1799	KSAQISCTAENKAGTASRDFFIQNIAPTFKNEGDOETIFRESEITILDCPVSLGDFQIT	1858
Qy	252	W-----T-----KD-----255	
Db	1859	WMKQGLPLTENDAIFTLDNTRLILNANRDHEDIYTCVANNTAGQVSKDFVVVQVLPKI	1918
Qy	256	-----GN-----L-----A-----S-----260	
Db	1919	KNAVTLNEGEIEIILTCDAENPTPTAKDFNQDLPKEAVFVNNHTVVVNNVTKYH	1978
Qy	261	T-----L-----GQ-----H-----K-----266	
Db	1979	TGVYKCVATNKVQAVKTIINVHVRTPREFSGLTSELTAVNLTFRSITLECDVDAIGVI	2038
Qy	267	-----G-P-----I-----F-----270	
Db	2039	SWTVNGKPFLEAETDGVQTLAGGRFLHIVSAKTDHGSYACTVNEAGVATKTNLFVQVP	2098
Qy	271	-----AL-----K-----W-----274	
Db	2099	PTIVNEGGEVTVIENNSLVLPCEVTGKPNPVVTWKDGRPVGLKSVQVLSGEQGFVH	2158
Qy	275	-----N-----K-----KG-----278	
Db	2159	AEIAHKSVCIMAKNDVGTAEISFDVDTIIFRPMIOKIXINVTIAGGALPFKCPIDDDK	2218
Qy	279	NF-----IL-----S-----AG-----285	
Db	2219	NFKGQIILWNYOPIDLEADARITLSDNRRLTILNVNTEDEGOYSCRKNDAGENSFD	2278
Qy	286	-----V-----D-----KT-----TI-----I-W--D--A-H-----296	
Db	2279	FKATVLVPPPTIIMLDKDKKTAVEHSTVTLSCPATGKPEPDITWFKDGEAIIHENIADII	2338
Qy	297	-----T-----G-----EA-----K-----Q-302	
Db	2339	PNGELANGNOLKTRIKEGDAGKYTCESADNSAGSVEQVNVNVTIPKIEKDGIPSDYESQ	2398
Qy	303	Q-----303	
Db	2399	QNERVVISCPVYARPPAKITWLKAGKPLQSDKFKVKTSAHQKLVLFKLRDSSKYTCIA	2458

QY 304 -----F-----P-F-----H-----S----- 308
Db 2459 TNEAGTKDRDFKVSMLVAPSFDEPNIVRRITVNSGNPSTLHCPAKGSPSTITWLKDGNA 2518
QY 309 -----A-----P----- 310
Db 2519 IEPNDRYVPFDAGRLQISKTEGSDQGRYTCIATNSVSGDDLENTLEVIIPPIDGERRE 2578
QY 311 --A-----L-D-----VD--WO-----S-N----- 319
Db 2579 AVAVIEGFSSELFCDSNSTGVDEWQXGDLTINQDTLGRDSFTIQIPSSGKMSFLSARKS 2638
QY 320 -----N-----T-----FA-----S-----C-S-T----- 327
Db 2639 DSGRYTCIVRNPAGEARKLPDFAVNDPPSISDELSSANIQTIVPYYPVEINCVVSGSPHP 2698
QY 328 -----DM-----C-----I- 331
Db 2699 KVTYLFDDKPLEPDSAAYELTNNGETLKIVRSQVEHAGTYTCEAONNVGKARKDFLVRVT 2758
QY 332 --H-----V-----C-----KL-----G----- 337
Db 2759 APHPFKEREVARVGDTHMLTCNAESSVPLSSVYVWHAHDESVOQGVITSKYAANEKTL 2818
QY 338 --Q-----Q-----DR-----PI-----K----- 343
Db 2819 NVTNIQDDBGFFYCTAVNEAGITKKFFKLVIVETPYFLDQOKLYPIILGKRLTLDOSAT 2878
QY 344 --T-----F----- 345
Db 2879 GTPPTILFMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGREKDMVVE 2938
QY 346 -----Q-----G-H----- 348
Db 2939 VLLPPKLSKEWINEVVOAGDPLFLECPIDTSGVHITWSRQFGKOGOLDMRAQSSDKSK 2998
QY 349 --T-----N-----E-----VN-----A----- 354
Db 2999 LYIMQATPEDADSYCIAVNDAGAEAVQVTVNTPPKIFGDSFSTTEIVADTTLPIPCR 3058
QY 355 -----IKW----- 358
Db 3059 TEGIPPEISWFLDKPILFMPGVTYKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQD 3118
QY 359 -----P-----T-----GN----- 362
Db 3119 TYVEISEPPRVNMASEVMRVVEGRQTTIRCEVFNPEPVVNNLKGEPYTSLLQFSTKL 3178
QY 363 --L-----LA-----SC-----S-----D-----D-----M----- 371
Db 3179 SYLHLRETTLDGTYTCIATNKAGESOTTDDVEVLVPPRIEDEERVLOKSGNTYVHC 3238
QY 372 ----- 373
Db 3239 QVTGRVPVYVTKRNGKIEQFNPVLHNRATRADEGKYSCIASNEAGTAVADFLIDVFT 3298
QY 374 -----KI-----W-----S-----M-K 379
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMDNIILSPRGDTLMILK 3358
QY 380 -----QD-----N-----C----- 383
Db 3359 AQRFDGGLYTCVATNSYGSQDFKVNVTYKPYIDETIDOTPKAVAGGEIILKCPVLGNP 3418
QY 384 -----Y-----H-----DL-----Q-----A-----H-NKE-I 394
Db 3419 TPTVTWKRGDDAVPNDNRHTIVNNYDLKINSVTTEDAGQVSCIAVNEAGNLITHYAAEVI 3478
QY 395 -----Y-----TI-K-----W-----SP----- 401
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDGSHIT 3538

QY 402 -----T----- 402
Db 3539 INKAKLSDGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLECPI 3598
QY 403 -G--P-----G-----T-----NN----- 408
Db 3599 SGIPQPDVITWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGKASHD 3658
QY 409 -----P----- 409
Db 3659 FSLDVLSPPEFDHGTQPTIKREGDTITLTCPKLAEDIADQVMDVSWTKDSRALDGLT 3718
QY 410 -N-----A--N--L-----ML--A-SAS--F-----D----- 421
Db 3719 DNVDISDDGKRLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPPIDISRNVDVQPQ 3778
QY 422 -----S-----T-----VR-----L-----W-- 427
Db 3779 VAVNQPTIMRCAVTHGHPFSIKWLKNGKEVTDDENIRIVEQGVQLILRTDSDHAGKWSC 3838
QY 428 -----DV-----D--R--G--I--C----- 434
Db 3839 VAENDAGVKELEWLVDFTPPVSVKSDNPICALGETITLFCVASGNPYPOLKWKAGGSL 3898
QY 435 -----I-H-----T----- 437
Db 3899 IFDSPDGARISLKGARLDIPHLKKTVDVGTQCALNAAGTSEASVSDVLVPPPEINRDOI 3958
QY 438 --LT-----K-----H-----Q----- 442
Db 3959 DMSRPLPAQOSLTLOCLAQKVPQMRWTLNGTALTHTSPGITVASDSTFIQINNVSLS 4018
QY 443 -----E-----PV-----Y-----S-- 447
Db 4019 KGVTYCAENVAGSDNLMYNDVVOQAPVISNGGTQKQVIEGELAVIECLVEGYPAPQVSWL 4078
QY 448 -----V-----A----- 449
Db 4079 RNGNRVETGVQVRYVTDGRMLTIIEARSLDSGIVLCSATNEAGSAQAQVTLVLSVPKI 4138
QY 450 -----FS-----P-----DG----- 454
Db 4139 ITSTPGVLTPSSGKSPFLPCAVRGYPDPIISWTLNGNDIKOGENGHTIGADGTLHIEAE 4198
QY 455 -R--Y-----L-----A-----SG----- 460
Db 4199 ERHLIYECTAKNDAGADTLEFPVQTVIVAPKISTSGNRYNGSEGTETVIKCEIESSESEF 4258
QY 461 -----S--F--D--K-----C----- 465
Db 4259 SMSKNGVPLLPNNLIFSEDYKLIKILSTRLSDQGSYSCTAANKAGNATQKTNLNVGAP 4318
QY 466 -----VH-----I-W-----NT----- 471
Db 4319 KIMERPTQVHKGDQVTLWCEASGVPPAITYKDNELLTNTGTVDDETATTKKSVIFSS 4378
QY 472 -----TG-----Q----- 474
Db 4379 ISPSQAGVYTCKAENWVASTEDIDLIVMIPPEVVPERNVNTNPRQTVFLSCNATGIPE 4438
QY 475 -----A-----L-----V-----H----- 478
Db 4439 PVISWNRDSNIAIQNNEKYQILGTLAIRNVLPPDDGFYHCIAKSADAGQIATKRLIVNK 4498
QY 479 -----S--YR--G-- 482
Db 4499 PSDRPAPIWVECEKGPKKTEYIMIDRGDTPDDNPOLLPWKQVEDSSLSNGSIAYRCMPGP 4558
QY 483 -----T-----G-----G--I-----FE-- 488
Db 4559 RSSRTVLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPPHPTITWAKDGKLEDKFEITA 4618
QY 489 -----V--C-- 490

Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTYSAFINVDNNILPTPKSSNQKNVAVITCYE 4678
QY 491 -----W-----N-A----- 493
Db 4679 RNOAYSRLTWEYNGVPMPLAGIHPMNGSLVILDTSSLKGBLELYTCKVRNRRHS 4738
QY 494 -----A-G-----DKV-----G-----AS-----A--- 502
Db 4739 IPHLTSAFEGVPEVKTIDRKEVNGDSVLVDCEVTSPLTHVVTYKNDQKMLDDDAIYV 4798
QY 503 -----S-----D-GSV-CV-----LD-----L-----R-----K 514
Db 4799 LPNNSLVLLNVEKYDEG-VYKVASNSIGKAFDDTQLNVYGGSSREAYK 4847

RESULT 2
T43290
hemictin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T43290; T20993; T24734
R:Vogel, B.E.; Hedgecock, E.M.
A:Description: Hemictin is required for hemidesmosome mediated cell adhesion and germ-
A:Reference number: 222396
A:Accession: T43290
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5198 <W0>
A:Cross-references: UNIPROT:O76518; EMBL:AF074501; PIDN:AAC26792.1
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20993
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W1>
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CBSP:F15G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24734
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W2>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CBSP:F15G9.4b
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: him-4; F15G9.4b
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
A:2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/1

Query Match
Best Local Similarity 9.2%; Pred. No. 6.6e-53;
Matches 447; Conservative 52; Mismatches 12; Indels 4362; Gaps 387;

QY 1 M-----SI-----SS-----D-----E----- 7
Db 1 MGRSPSWLYGVGLLLATTCSSVNDKNDKNDPTCKSLAFVFDITGSMFDDLVQVREGAAK 60
QY 8 -----V-----N-----F-----L-----VY----- 13
Db 61 IFKTVMAQREKLIYNIWPFHPDYLGEIINTDSTYFMRLSKVYVHGGDCPEKTLTG 120
QY 14 -----R-----Y-----L-----QE-----SG-----F-----SHSAF-T 27
Db 121 ILKALQISLSPSIYVFTDARSKDYLHLEDEVLNTIQEKSSVYVFMVMTGCGNETHPGFRT 180
QY 28 -----FG-I-----E-----S-----H-ISQ-----S-NI----- 39

Db 181 YEKIAAASFGQVPHLEKSDVSTVLEVRHAVQKKVHLMAYEARGGTYSRNPVDDKHL 240
QY 40 -----N-----GA-----LV-----P-P----- 46
Db 241 ELTISLSDGKDDNDLDIVLRDPEGRTVDKRLYSKEGGTIDLKNVKLIRLKDPSGVWTV 300
QY 47 -----A-----A-----A-----LI-----S-I----- 52
Db 301 NNTSRUKHTIRVFGHGAVDFKYGFASRPLDRDRIELARPRVPLVNDQTVLLINMTGLIPPGTV 360
QY 53 --I-----Q-----KGL-----Q-Y-----VE-----AE----- 63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPPVPKGLFFVRVQGVQDNYEFMRAPT 420
QY 64 -----V-----S-I-----N-----ED-----GTLF----- 73
Db 421 AIGSVIVGGPRAPMSPIHQEFVGRDLNLSCTVESASAYTIYWKTEGDIIGPLFYHNTD 480
QY 74 -----D-----Q-----R-----P-I-----E----- 79
Db 481 TSVWTTPELSLKDAGEYECRVISNNGVSVKTRVETRESPPPEIFGVNRVSVPLGEAFLH 540
QY 80 -----S-----L-----SL-I-----DA----- 86
Db 541 CSTRSAGEVEIRWTRYGATVFNCPNTERNPTNGTLKIHVTRADAGVYECMARNAGMST 600
QY 87 -----VM-----P-DV----- 92
Db 601 RKMRLDIMEPPSVKVTPODVYFNMREGVNLSCAMGDPKPEVHWYFKGRHLLNDYKYQVG 660
QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q----- 105
Db 661 QDSKFLYIRDATHDEGTECRAMSQGA-RDTTDLMLATPPKVEIIONKMVGGRDRV 719
QY 106 -----Q-H-----A-----AAA-----AA----- 113
Db 720 SPECKTIRKPHPKIRWFKNGDLIKPDDYIKINEGQLHMGAKDEADAGAYSCVGENMAG 779
QY 114 -----AAAA-----T-----N-----Q-----Q-GS-----A-KN----- 126
Db 780 KDQVQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEMQKGNVLLATLNNP 839
QY 127 -----GEN-----T-----A-N----- 132
Db 840 RYTQLADG-NLLITDAQIEDQOFTCIARTYQQSQSTTLMTVGLVSPVLGHVHPPEEQ 898
QY 133 -----G-----EE-----N-----G----- 137
Db 899 IEGQDLTLCVVVGLTPKPSIVWIKDDKPVIEGPTIKIEGGSLRLRGGNPKDEGKYTC 958
QY 138 -----A-----H-----TI-----A-N-NH----- 145
Db 959 IAVSPAGNSTLAINVOLIKKPEFVYKPEGGIVFKPTISGMDEKHAVVNSVTHDVLDEGF 1018
QY 146 -----T----- 146
Db 1019 AIPCVSVGTPPPIITWYLDGRPIITPNSRDFVTADNTLIVRKADKSYSGVYTCAINSAG 1078
QY 147 D-----M-----E-----VD-----GD-----V----- 155
Db 1079 DNEQKTTIRIMNTPMISPGQSSFNMVVDLFTIPCDYGDPKPVTIWLDDDKPFTGEGVN 1138
QY 156 E-----I-----PS-----N-K-A-----VVL----- 165
Db 1139 EDGSLTIPNVNEAHRGTTTCHAQNAAGNDTRVTLTPTTINAENQEKIALQNDIVL 1198
QY 166 -----R----- 167
Db 1199 ECPAKALPPVRLWTEGKIDSQIPHTIREDGALVQLONVKLENTGVFCVQVSNLAGED 1258
QY 168 -----HE-----SEV-----F-----I-C-A-----WNP-----V----- 180

Db 1259 SLSYTLTWHKPKIIEBPGVVDVVKGFTEIEIPCRATGVPEVIRTNKNGIDLKMDEKKF 1318
QY 181 S-D-L-L-V-S--G--SG--D--S--TA-- 193
Db 1319 SVDNLGTLRIEADKNDIGNYCVVNEAGTSOMTHVDVQBPPIILPSTQTNNTAVVGD 1378
QY 194 RI--W--NL--S-ENST--S-- 204
Db 1379 RVELKCVASPPASVTFRRGIAIGTDTKGVVSDGTLVIQASVEDATIYCKASNP 1438
QY 205 -G--P--TQ--L--V--L-- 211
Db 1439 AGRAEANLQVTVIASPDIKDPDVVTQESIKESHPPSLYCFVFNPLFOISWYLNDRKPLID 1498
QY 212 --R-H--CI-R-- 216
Db 1499 DKTSWTSDDKRLHVFKAKITDSGVYKCVARNAAGEGSKSFQEVIVPLNLDESXYKKK 1558
QY 217 --EG--G--Q--D--V-P--SN-- 225
Db 1559 VFAKEGEVTLGCPVSGFPVPPQINWVDGTVVEFGKYGKATLSNDGLTLHFDVSVMKQE 1618
QY 226 --KQVT-S-L--D--W 233
Db 1619 GNYHCVAQSGKNILDIDVLSVLAVPIVGEDDNLVFLGKDLSLSCDLQTESDDKTTFW 1678
QY 234 --N-SE--G--G--TL-L-- 240
Db 1679 SINGESDRPDVQIIPSDGHRLYITDAKPENNGKYMCRVNTNSAGKAERTLTLVLEPPVF 1738
QY 241 --A--TG--SY-D-- 246
Db 1739 VEPVFEANQKLGNNPIILCQVGTGNPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLGT 1798
QY 247 --G--F--A--R--I-- 251
Db 1799 KSAQISCTAENKAGTASRDFQIONAAPTFRKNEGDQETIFRESEITITLDCPVSLGDFQIT 1858
QY 252 W--KO--T-- 255
Db 1859 WMKQGLPLTENDAIPTLNDTRLTLNANRDHEDIYTCVANNTAGQVSKDFVQVVLPKI 1918
QY 256 --GN--L-A--S-- 260
Db 1919 KNAVVTLEINEGEEIILTCDAENPPTAKWDNFQGDLPKEAVFVNNHTVVNNVTKYH 1978
QY 261 T--L--CQ--H-K-- 266
Db 1979 TGVIKCYATNKVQAVKTINVHVRTKPRFESGLTESELTVNLFRSITLBCDVEDDAIGVGI 2038
QY 267 --G-P--I--F-- 270
Db 2039 SWTVNGKPFLEATDGVQTLAGGRFLHIVSAKTDHGSYACTVNEAGVATKTNLQVVP 2098
QY 271 --AL--K--W-- 274
Db 2099 PTIVNEGGEVTVIENSLVLPCEVTKGNPVVVTWKDGRPVGLKSVQVLSGQPFKIVH 2158
QY 275 --N--K--KG-- 278
Db 2159 AEIAHRGSYICMAKNDVGTAEISFDVDIITRPMIOKGIKNIVTAIKGGALPFRKPIDDDK 2218
QY 279 NF--IL--S--AG-- 285
Db 2219 NFKGQIILWNYOPIDLEADARITLSNDRRLTLNVTEDEGQYSCRVDNAGNSFD 2278
QY 286 --V--D--KT--TI--I-W-D--A-H-- 296
Db 2279 FKATLVPPPIIMLDKDKNKAVEHSTVILSCPATGKPEPDIWFKDGEAIIHENIADI 2338
QY 297 --T--G--BA--K--Q 302
Db 2339 PNGELNGNQLKIRIKEGDAKGKYTEADNSAGSVEQDVNNVNTIPIKIEKGIPSDYESQ 2398

QY 303 Q-----F-- 304
Db 2399 QNERVVISCPVYARPPAKITWLKAGKPLQSDKFKVTSANGOKLYLFLKRETDSSKYTCIA 2458
QY 305 -----P-F-----H-----S----- 308
Db 2459 TNEAGTKRDKFKVSMVLVAFSPDEPNIVRRITVNSGNPSTLHCPAKGSPSPITWLKDGNA 2518
QY 309 -----A-----P----- 310
Db 2519 IEPNDRYVFDAGRQLQISKTBSGQGYTCIATNSVGSDDLLENTLEVIIPPIDGERRE 2578
QY 311 --A--L-D--VD--WO--S-N----- 319
Db 2579 AVAVIEGFSSELCFCDNSNSTGVDVWQKDGTLINQDTLRGDSFIQIPSSGKMSFLSARKS 2638
QY 320 -----N--T--FA-----S-----C-S-T-- 327
Db 2639 DSGRYTCIVRNPAAGEARKLFDFAVNDPPSISDELSSANIQTIVPYYPVEINCVSVPSPHP 2698
QY 328 -----DM--C-----I- 331
Db 2699 KYVWLFDDKPLBPDSSAAYELTNGETLKIVRQVEHAGTYTCEAQNNGVKARKDFLVRVT 2758
QY 332 --H--V--C--KL--G-- 337
Db 2759 APHPFKEREVEVVARVGDTMLTLCNAESSVPLSSVYVWHAHDESVQNGVITSKYAANEKTL 2818
QY 338 --Q--DR--PI--K-- 343
Db 2819 NVTNIQLDDGFGYCTAVNEAGITTKFKFLIVETPYFLDQOKLYPIILGKRLTDCSAT 2878
QY 344 --T--F----- 345
Db 2879 GTPPPILFMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVE 2938
QY 346 --Q--G-H----- 348
Db 2939 VLLPPKLSKEWINVEVQAGDPLTECPIEDTSGVHITWSRQFGKQGLDMRAQSSSDSKS 2998
QY 349 --T--N--E--VN-----A----- 354
Db 2999 LVIMQATPEDADSYSCIAVDAGAEAVFQVTVNTPPKIFGDSFSTTEIVADTLEIPCR 3058
QY 355 -----IKW-----D 358
Db 3059 TEGIPPEISWFLDGKPILEMPGVTYKQGLSLRIDNIKNQEGRYTCVAENKAGRAEQD 3118
QY 359 -----P--T--GN-----L----- 362
Db 3119 TYVEISEPVRVMASEVMRVESGRQTTIRCEVFGNPEPVVNWLDKGEPTYSDLLQFSTKL 3178
QY 363 --L--LA--SC--D--D-----M-- 371
Db 3179 SYLHLRETTLAGGTVTCIATKNAGSQTTTDEVLVPPRIEDEERVLOGKEGNTVMVHC 3238
QY 372 -----T-----L----- 373
Db 3239 QVTGRPVYVTKWRNGKEIEQPNVPLHNRATRADEGKYSCIASNEAGTAVADFLIDVFT 3298
QY 374 -----KI--W-----S-----M--K 379
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPIISMLKGRPFNMWNIILSPRGDITMLK 3358
QY 380 -----QD--N-----C----- 383
Db 3359 AQRFDGGLYTCVATNSYGDSEQDFKNVYVTKPIDETIDQTPKAVAGGBIILKCPVLGNP 3418
QY 384 -----V--H--DL-----Q-----A-----H-NKE-I 394
Db 3419 TPTVTWKRGDDAVPNDRSRTIIVNNDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVI 3478

QY 395 -----Y--TI-K-----W-----SP-----401
Db 3479 GKPTVRKGNLYEVIENDTTIMDCGVTSRPLPSISWFRGDKPVLYLDRYSISPDGSHIT 3538
QY 402 -----T-----402
Db 3539 INKAKLSDGKVICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYLCPI 3598
QY 403 -G--P-----G--T-----NN-----408
Db 3599 SGIPQPDVITKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGKASHD 3658
QY 409 -----P-----409
Db 3659 PSLDVLSPPEDIHGTOPTIKREGDTITLCPIKLAEDIAQVMDVSWTKDSRALDGLT 3718
QY 410 -N-----A--N--L-----ML--A--SAS--F-----D-----421
Db 3719 DNVDISDDGKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPVIDISRNDVQPO 3778
QY 422 -----S-----T-----VR-----L-----W-----427
Db 3779 VAVNQPTIMRCVTHGTFPPSKLWLNKKEVTDENIRIVEQGVQLILRTDSHAGKWC 3838
QY 428 -----DV-----D--R--G--I--C-----434
Db 3839 VAENDAGKLEWLDVFTPPVSVKSDNPICALGETITILFCNASGNYPQLKWKAGGSL 3898
QY 435 -----I--H-----T-----437
Db 3899 IPDSPDGARISLKGARLDIPLHKKTVDGYTCQALNAAGTSEASVSDVLVPPPEINRDI 3958
QY 438 -----LT-----K-----H-----Q-----442
Db 3959 DMSRLPAQOSLTILQCLAQKQVPQMWTLNGTALTHTPGITVASDSTPIQINNVSLS 4018
QY 443 -----E-----PV-----Y-----S--447
Db 4019 KGVVTCVAENVAGSDNLMNVVDVQAPVINSNGGTVKQVIEGELAVIECLVEGYAPQVSWL 4078
QY 448 -----V-----A-----449
Db 4079 RGNRNVETGVGVRYVTDGRMLTIIARSILDSGLYLCSATNAGSAQAQVTLVLSPKI 4138
QY 450 -----PS-----P-----DG-----454
Db 4139 ITSTPGVLTSPSSGKSLPCAVRGYPDPPIISWTNLNGNDIKDNGHTIGADGTLHIEKAE 4198
QY 455 -R--Y-----L-----A--SG-----460
Db 4199 ERHLIVECTAKNDAGADTLEFPVQTTIVAPKISTSGNRYINGSGTETVIKCEIESSESEF 4258
QY 461 -----S--F--D--K-----C-----465
Db 4259 SMSKNGVPLPNSNLIIFSEDKILKILSTRLSDQGEYSCTAANKAGNATQKTNLVGVAP 4318
QY 466 -----VH-----I--W-----NT-----471
Db 4319 KIMERPTQVHKGDQVTLWCEASGVPOPAITWYKDNELLTNTGVDETATTKKSVIFSS 4378
QY 472 -----Q-----TG-----474
Db 4379 ISPSQAGVYTCKAENVWASTEEDIDLIVIMPEWPMNVSTNPQTQVFLSCNATGPE 4438
QY 475 -----A-----L--V-----H-----478
Db 4439 PVISWRDSNIALONNEKYQILGTLTAIRNLVLPDDGFFHCTAKSDAGOKIATRLIVNK 4498
QY 479 -----S--YR--G--482
Db 4499 PSRPAPIWVECDKGPKEWIDRGDTPDNPQLLPWKVDSDSLNGSIAYRCMPGP 4558
QY 483 -----T-----G-----G--I-----FE--488

Db 4559 RSSRTVLLHAAPOFIIVKPKNTTAAIGAIVELRCSAAGPPHTTITWAKDGKLIEDSKPEIA 4618
QY 489 -----V--C--490
Db 4619 YSHLKYTLNSTSDSGEYTCMAQNSVGSSTVSFAFVNDNNILPTPKPSSNKNQNAVITCYE 4678
QY 491 -----W-----N--A-----493
Db 4679 RNOAYSRLGTWEYNGVMPKPNLAGIHFMNNGSLVILDTSSLKEGDLELYTCKVNRNRHS 4738
QY 494 -----A--G-----DKV-----G-----AS-----A--502
Db 4739 IPHLTSAFEGVPEVKTIQKVENNGSVVLDCVETSDPLTTHVVVTKNDQKMLDDDAIYV 4798
QY 503 -----S-----D--GSV--CV-----L--D--L-----512
Db 4799 LPNNSLVLLNVEKYDEG--VYKCVASNSIGKAPDDTQLVNVEGDFLPTGTGEGSGINIDDS 4857
QY 513 -----R--K 514
Db 4858 SNAGSSRREAYK 4870

RESULT 3

I38346

elastic titin - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004

C;Accession: I38346

R;Label: S.; Kolmerer, B.

Science 270, 293-295, 1995

A;Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.

A;Reference number: A57430; MUID:96026330; PMID:7569978

A;Accession: I38346

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-7962 <RES>

A;Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PID:CAA62189.1; PID:g101

C;Genetics:

A;Gene: GDB:TTN

A;Cross-references: GDB:127867; OMIM:188840

A;Map position: 2q31-2q31

Query Match

Best Local Similarity 71.0%; Score 2603.7; DB 2; Length 7962;

Matches 448; Conservative 52; Mismatches 11; Indels 4663; Gaps 377;

QY 1 MS-----IS--SDE--V-N-----9
Db 57 MTQPEDTYQLEIABAYPEDEGTYTFVANNVQVSTANLSLEAPESILHERIEQIEME 116
QY 10 -----F-----L--VY-----R--Y--15
Db 117 MKAAPVKKIKPIEVLALHGLAKFTCEIQSAPNRFQWFKAGREIVESDKSIRSKYIS 176
QY 16 ---L--Q-----E-----S--G--P--21
Db 177 SLEILRTQVDCGEYTCASNBSYGSVCTATLTVTEAYPPTFLSRPKSLTTFVGAARFI 236
QY 22 -----S-----H-----S--24
Db 237 CTVTGTPTVETIWKDGAALSPNWRISDAENKHILSNLTIOQRGVYSCASNKFGA 296
QY 25 ---A-----F-----T-----27
Db 297 DICQAEIIIDKPHFIKELEPVQSAINKVKHLECOVDEDRKVTVTWSKDGKLPFGKDYK 356
QY 28 -F-----G-----I-----30
Db 357 ICFEDIATLEIPLAKLKDQSGTYVCTASNEAGSSCSCTVTVREPPSFVKKVDPSYMLP 416
QY 31 -ES--H-----ISQ-----SN-----G--41

Db	417	GESARLHCKLKGSPVI-QVTFWFKNNKELSESNTRVMYFVNSEAILDITDVKVEDSGSYSC	475
Qy	42	-A- - - - - LV - - - - - P - - - - - PA - - - - - AL - - - - - IS - - - - -	51
Db	476	EAVNDVSGSDSCSTEIVIKEPSPFIKTLEPAIDIVRGTHALLOCEVSGTGPFESWFKDKQKQ	535
Qy	52	- - - - - I - - - - -	52
Db	536	IRSSKKYRLFQSKSLVCLIEIFSNFNSADVGEYECVAVNEVGKCGCMATHLLKEPPTFKVKV	595
Qy	53	-I- - - - - Q - - - - - KG - - - - - L - - - - - Q - - - - -	58
Db	596	DDLIALGGQTVTLQAARVGSPI SVTWKMGQEVIREDGKIRMSFNGVAVLLIIPDVQISF	655
Qy	59	-Y- - - - - VB- AE- - - - -	63
Db	656	GKTYTCLAEAGSQTSGVELIVKEPAKIIERELIQVTAGDPATLEYTVAGTPELKPWK	715
Qy	64	- - - - - V- S - - - - - I - - - - - N - - - - - E- D - - - - - G - - - - - T- L - - - - -	72
Db	716	YKGRPLVASKKYRISPKNVNAQLKFYSAEHLDSGQVTFEISNEVGSSCETTFTVLDRD	775
Qy	73	-F- - - - - D - - - - - G - - - - - R - - - - - P - - - - - I - - - - - E - - - - - SL 81	
Db	776	IAPFTFKPLRNVDSVNGTCLRDCKIAGSLPMRVSWFKDGKEIAASDRYIAFVEGTASL	835
Qy	82	SLI - - - - - DA - - - - - RATNSVGSKDSSGALLIVQBPSPFVTKPGSKDVLPGSAVCLKSTFQ	895
Db	836	EIIIRVMDNAGNFTCRATNSVGSKDSSGALLIVQBPSPFVTKPGSKDVLPGSAVCLKSTFQ	895
Qy	94	- - - - - T- R - - - - -	95
Db	896	GSTPLTIRWFKGNKELVGGSCYITKEALESLLYLVTKSDSGTYTCKYSNVAGGVECS	955
Qy	96	- - - - - Q - - - - - Q- A - - - - - Y - - - - -	99
Db	956	ANLFVKEPATFEVKLEPSQLLKGDATQACKVTGTPPKITWPANDREIKESKHRMSF	1015
Qy	100	- - - - - R - - - - - D - - - - - KL - - - - - AQ - - - - -	105
Db	1016	VESTAVILRTDVIGEDSGEYMCQAQNEAGSDHCSSIVIVKESPYTFKEFPIELVKEYDV	1075
Qy	106	- - - - - AA - - - - - Q- H - - - - - AA- A - - - - -	110
Db	1076	MLLAEVACTPPFEITWPKDNTILRSGRKYKTFIQLHLVSLQLKFVAADAGEYQCRVTNE	1135
Qy	111	- - - - - AA - - - - - AAA- AAT - - - - -	118
Db	1136	VGSSICARVTLRPPSPFIKKIESTSSLRGGTAFAQATLKGLPITVTWLKDSDEITDD	1195
Qy	119	- - - - - N - - - - - Q - - - - -	120
Db	1196	NIRMTFENNVAASLYSGIEVKHGKYVCOAKNDAGIQRCSALLSVKEPATITBEAVSIDV	1255
Qy	121	-QG- - - - - S - - - - - AK - - - - -	125
Db	1256	TOGDPAITQVKSFTKEITAKWPKDGOBELTGLSKYKISVTTDTVSLIKIISTEKKDSGEYT	1315
Qy	126	- - - - - N - - - - - N - - - - - G - - - - -	127
Db	1316	FEVQNDVRSSCKARINVLDLIIPPSFTKKLKKMDSIKGSPIDLECIVAGSHPSIQWPK	1375
Qy	128	- - - - - E - - - - - NTA - - - - - N - - - - - G - - - - - E - - - - -	134
Db	1376	DDQEISASEKYKFSFHONTAFLEISOLEGTDSGTYTCSATNKAGHNQCSGHLTVKEPPYF	1435
Qy	135	-E- - - - - N - - - - - G- A - - - - - H - - - - -	139
Db	1436	VEKPSQODVNPNTRVQLKALVGGTAPMTIKWFKDNKELHGAARSVWKDDTSTSLLELFAA	1495
Qy	140	- - - - - T- I - - - - - A - - - - - N - - - - -	143

Db	1496	KATDSGYI	COLSNDVGTATSKATLFVKEBPQFIKKPSPVLVRNGQSTTFECQITGTTPK	1555
Qy	144	-----N-----H-----T-----D-----	-----	147
Db	1556	IRVSWYLDGNEITAIQKHGISFIDGLATFOISGARVENSGTYVCEARNdagTASCSIELK	1615	
Qy	148	-----MM-EV-----D-----G-----	D-----V-----	155
Db	1616	VKEPPTFIRELKPVEVVKYSDVELECEVTGTPPFEVTLKNNREIRSSKKYTLTDREVSVF	1675	
Qy	156	-----E-I-----J-----PS-----N-----K-----	-----	161
Db	1676	NLHITKCDPSDTGEYQCIIVSNEGSCSCSTRVALKEPPSPFIKKIENTTTTVLKSSATFQST	1735	
Qy	162	A-----V-V-L-R-----GH-----	ES-----	170
Db	1736	VAGSPISITWLKDDQILDDEDDNVIYSFVDSVATLQIRSDVNGHSGRYTCQAKNEGVER	1795	
Qy	171	-----	-----	170
Db	1796	CYAFLLVQEPQIVKAKSVDTKDPMTLECVVAGTPELKVKWKDKGQIVPSRYFSMS	1855	
Qy	171	-----E-----V-----P-----	-----	173
Db	1856	FENNVSFRIQSVMKODSGOYTFKVENDFGSSCDAYLRVLDQNIIPPSFTKKLTKMDKVL	1915	
Qy	174	---I---C---A-W-----	-----	177
Db	1916	GSSIHECKVSGSLPISAQWFKDGKEISTAKYRLVCHERSVSLEVNLEBDTANYTCK	1975	
Qy	178	-N-----P---V-----S-----DL-L 184	-----	184
Db	1976	VSNAVGDACSGILTVEKPPSFLVKPGRQQAIPDSTVEFKAILKGTPPPKIKWFKDDVEL	2035	
Qy	185	VSG-----S-----G-DS-----T-----	-----	192
Db	2036	VSGPKCFIGLEGSTSFNLYSVDASKTGQYCHVTNDVGSDSCTTLLVTEPPKPFVKKLE	2095	
Qy	193	---A---R-----I---W-----NLS--- 199	-----	199
Db	2096	ASKIVAGDSRLLECKIAGSPERVVWFRNEHELPAADKYRMTFIDSVAVIQNNLISTED	2155	
Qy	200	---E-N---ST-----VL-----RHC----- 214	-----	214
Db	2156	SGDFICEAQNAGTSCSTKVIKVEPPVFSSPPPIVETLKNABVSLECELSGTPPEVVM	2215	
Qy	208	-QL-----	-----	214
Db	2216	YDKRQLRSSKKYKIASKNPHTSIHILNVDTSDIGYHCKAQNVEVSDTCVCTVKLKEPP	2275	
Qy	215	-----IRE-----	-----	217
Db	2276	RFVSKNSLTVVAGEPAELQASIEGAQPIFVQWLKEKEVIRESENIITFVENVATLQF	2335	
Qy	218	-----GG-----	-----	219
Db	2336	AKAEPANAGKYICQIKNDGGRENMATLMLVPEPAVIVEKAGPMVTTVGETCTLECKVAGT	2395	
Qy	220	-----QD-----	-----	221
Db	2396	PELSVEWYKDGKLLTSQKHKFSFYNKISSRLILSVEROAGTYTFQVQNNVKSCTAV	2455	
Qy	222	-----V-PS-----N-----K----- 226	-----	226
Db	2456	VDVSDRAVPSPFTRLKNTGGVLGASCILECKVAGSSPISVAMFHEKTKIVSGAKYQTTF	2515	
Qy	227	-D-V-T---SLD-----	-----	232
Db	2516	SDNVCTQLNLSLSDSMGNVTCVAANNVAGSDECRVLTVQEPSPSFYKEPEPLEVLPGKV	2575	
Qy	233	-----W-----N-----S-E----- 236	-----	236
Db	2576	TFTSVIRGTPEPKVNFGRGARELVKGDRCNIVPEDTVALELFPNIDISOSGEYTCVVSNN	2635	

QY 237 -G---T-LL---A---TG---S---Y---DGF--- 248
Db 2636 AQASCTRLRVKBPAAFLKRLSDHSVEPGKSIILESTYTGTLPSVTWKXGDNITSE 2695
QY 249 ---A---A-R--- 250
Db 2696 KCVITTEKTCILBNLSTKRDAGQYSCIEAGRDVCGALVSTLBPYFVTEPLEA 2755
QY 251 ---I---W---TK---D-G--- 256
Db 2756 AVGSVSLQCVAGTPEITSWYRGDKLRTPEYRTYFTNNVATLVFNKVNINDSGEYT 2815
QY 257 ---N-L---AS---TL---G---Q--- 264
Db 2816 CKAENSGTASSKTVFRIQERQLPPSFARQLKIEQVGLPVTLTCLRLNGSAPIQVCWYR 2875
QY 265 ---H---K---G--- 267
Db 2876 DGVLLRDHENLQTSFVDNVATLKLQDLSHGQYSCSASNPLGTASSSARLTAREPKS 2935
QY 268 ---P---I---F---A---LK---W---NK--- 276
Db 2936 PFDIKPVSIDVIAGESADFECHVTGAQPMRITWSKDKNKEIRPGNYTITCVGNTPHLRI 2995
QY 277 ---KG--- 278
Db 2996 LKVGKDSGQYTCQATNDVGDKMCASQLSVKEPPKFVKLEASKVAKQGESIQLECKISG 3055
QY 279 ---N---FI---L---SA---GV-D--- 287
Db 3056 SPEIKVSWFRNDSSELHESWKNYNSFINSVALLITNEASAEADSGDYICEAHNGVGDASCST 3115
QY 288 ---K---T---T---IIW--- 293
Db 3116 ALTAKPPVTQKPSVPVGLKGSVDILOCEISGTPPFVVMVKDRQVRNKKFKITSKH 3175
QY 294 -D---A---H---H---KATNEVGS DTCSCSVKFKBPFPVKLS DTS TLIGDAVEL 3235
3176 FDTNLHLNLEASDVGEYHCKATNEVGS DTCSCSVKFKBPFPVKLS DTS TLIGDAVEL 3235
QY 300 -A---K---Q--- 303
Db 3236 RAIVEGFOPISVWMLKDRGEVIRESENTRISFDINIA TLQGSPEASNSGKYCIQKND 3295
QY 304 ---F---P--- 306
Db 3296 GMBECASVLTVLEBPARIIEKPSPMVTVTGNPFALCECVVTGTPBELSAKWFQKGRSLSADSK 3355
QY 307 -H-S---A---P-A---L---D-V--- 315
Db 3356 HHITFINKVASLKI PCAEMSDKGLYSEFVNKNSVKNCTVSVHVS DRIVPPSPFIRKLKDV 3415
QY 316 ---W---QS---N--- 319
Db 3416 NAILGASVWLECRVSGSAPISVGVFQDGBEIVSGPKQCSSENVCTLNLSLEPSDTGI 3475
QY 320 ---N---TP--- 322
Db 3476 YTCVAANVAGSDECSAVLTVOEPPSPFQTPDPSVEVLPGMSLTFTSVIRGTPPPKVKWFKG 3535
QY 323 ---ASC--- 325
Db 3536 SRELVPGESCNISLEDPVTELEFVQPLESGDYSCLVITNDAGSASCTH LFKVEPATFV 3595
QY 326 ---S---T---D---M--- 329
Db 3596 KRLADFSVETGSPVILEATYGTTPPISVMSIKDEYLISQSERCSITWTEKSTILEILEST 3655
QY 330 ---C---I---HV---CKL-G--- 337
Db 3656 IEDYQYSCLIENEAGODICEALVSVLEPPYFTEPLEHVEAVIGEPA TLQCKVDGTFEIR 3715

QY 338 ---Q---D--- 339
Db 3716 ISWYKHTKLRSAPAYKMQFNNAVSLVINKVDHSDVGEYSCKADNSVGAVASSAVLVIK 3775
QY 340 -R---P---I-KT---F---Q---G---H--- 348
Db 3776 ARKLPFPFARKKLVHETLGPVAFECRINGSPLQVSWYKDGVLKKDDANLOTSPFHVNV 3935
QY 349 ---T---N---EV--- 352
Db 3836 ATLQILQTDOSHIGQYNCASNPLGTASSAKLILSEHEVPPFDLKPVSVDLALGESGT 3995
QY 353 ---NA-IK-W-D---PTGN---L-L---A-SC-S--- 368
Db 3896 FKCHVTGTAPIKITWAKDNREIRPGNYKMTLVKVGKDGAGQYTCYASNIA 3955
QY 369 ---D---DM-T-L-KI---W---SMK-Q--- 380
Db 3956 GKDSCSAQLGVQEPFRFIKKLEPSRIVKQDEFTRYECKIGSGPEIKVLMYKDET-EIQES 4014
QY 381 ---D---N---C---V---H--- 385
Db 4015 SKFRMSFVDSVAVLEMHNL SVEDSGDYTCFAHNAGSASSTSLKVKEPPIPRKKPHPIE 4074
QY 386 ---D---LO---AH--- 390
Db 4075 TLKGADVHLECELQGTGPPFHVSWYKDRBLRSKGYKIMSENLFTSIHILNVDAADIGEY 4134
QY 391 ---N---KEI-Y-TI---K 398
Db 4135 QCKATNDVGS DTCVGSIALKAPPRFVKLSDISTVVGVEQLQTTIEGAEPISVWFKDK 4194
QY 399 ---W---S--- 400
Db 4195 GEIVRESDNITWISYSENIATLOFSRVEPANAGKYTCQIKNDAGMOBFCATLSVLEPATIV 4254
QY 401 ---P---TG---P---G-T---N---N--- 408
Db 4255 EXPESIKVTGDTCTLECTVAGTPELSTKWFQDKGKELTSDNKYKISFPFNKVSGLKINVA 4314
QY 409 P---N---A---N-L---M---L--- 415
Db 4315 PSDSGVYSFEVQNPVKGKSCSTASLQVSDRTVPSPFTRKLKETNGLSGSSVWMECKYVGS 4374
QY 416 -ASAS-F---DS--- 422
Db 4375 PISVSWFHEGNEISSGRKYQTTLTNTCALTVMNLEESDSGDYTCIATNMAGSDECSAPL 4434
QY 423 TVR---L---W--- 427
Db 4435 TVREPPSFVQKDPMDVLTGTVNFTTSIVKGTTPPSVSWFKGSSELVPGDRCNVSLDSV 4494
QY 428 ---DVD-R-G-ICI---H--- 436
Db 4495 AELELPDVTSSGSEYTCIVSNEAGKASCTHLYIKAPAKFVKRLNDYSIEKGKPLILEG 4554
QY 437 ---T--- 437
Db 4555 TFGTTPPISVTWKNGINVTSPQR CNITTEKSPILEIPSTVEDAGQYNCYNIENASGKD 4614
QY 438 ---L---TK--- 440
Db 4615 SCSAQIILLEPPYFVKQLEPVKVSVDASLQCLAGTPEIGVSWYKGTKLRTPTYK 4674
QY 441 H---QE---P--- 444
Db 4675 HFRNNVATLVFNQVINDSGEYICKAENSVEGVSASTFLTVOEQKLPPSPSRQLRDVQET 4734
QY 445 V-YSAF---S-P---DG---R---Y--- 456
Db 4735 VGLPVVFDCAISGSEFISVSWYKDKRPLKDSPNVQTSFLDNTATLNI FKTDRLAGQYSC 4794
QY 457 LAS---GS---FD---KCVH---I---W--- 469

147 QY -----D--M--E-----VD-----152
1993 Db SEPTIGQSUVAKNPPFBPGKGTPEAVDMDKOHDLVWRPPINDGGSPITGYVVEKREK 2052
153 QY G-D-V--EI-----PS--N--KAV-----163
2053 Db GTDKWIKGTEITPCLGCECKATVPTLNENCEYEFVRKAINAAGPGSPDASKPIITKPR 2112
164 QY -----VL-----R-----G-----H-E-----169
2113 Db KLAPTILDTKNIRTYNFKSGEPIFLDINISGEPAVDVTWNQNNKSQVTSFSHIENLPY 2172
170 QY -S-----E-----V-F--I-----C-----175
2173 Db NTKVINNNPERKOTGLYKISAHNFYQDQVEFQINIITKPKGPGPLEVSEVHKDCKLK 2232
176 QY -----A-----W--N-----PVSD-----LLV-----185
2293 Db ESEPLETGSIIAKDPFSVPTKPGVPEPTDWTANKVELAWPEPASDGGSPIOGYIVEVKD 2352
186 QY -----S-----G-----D-S--TA-----193
2353 Db KYSPLWEKALETSPPTATVOGLIEGNEYQFRVVALKNLSEPSDPSKIIFAKKRYIA 2412
194 QY -----R-----I-----W-----N-----L-198
2413 Db PKIDRRNLRLITSSGTALKLDANI TGQAPKPVKEWKL SNVHLSQGNVTIETPDYTKLV 2472
199 QY -----S-E-----N-----S-----GP-----206
2473 Db IRPQRTDSGYLVATNTSGKDSVLNVVITDKPSPNGPLQISDVHKEGCHLKWKRPS 2532
207 QY -----T-Q-----L-----V-----L 211
2533 Db HDGTPLEYFQIDKLEBETGCWIPSCRSSTAPQVDVTGLSPGNEYKFRVSAVNAEGSQPL 2592
212 QY -----R-----H-----C-I-----215
2593 Db VGDESIVARNPFDEPKPENLKATMDKOHVDLAWTPPVIDGGSPISCVIIERQDKYKGW 2652
216 QY -R-----EG-----218
2653 Db ERALDVPADOCKATI PDLVEGQTYKFRVSAVNAAGTGEPSDSTPPIIAKARNKPIIDRS 2712
219 QY -----GQ-----D-V-----P-----S-----N--K-----226
2713 Db SLVEVRKAGOSFTFDCKVSGEPAPQTKMLKKKEVYSKDNVKTVDYNTKLVNSATR 2772
227 QY -----D--VT-----S-----230
2773 Db SDSGIYTVFAENANGESADVKVTVIDKFPAPPNGPLKVDEINSECTLHWNPPDDGGQP 2832
231 QY -----LD-----W-----N-----SE-----236
2833 Db IDNVVVKLDETTGRWMTAGETGCPVTALKVGGLTPGHKYKFRVAKNRQGTSEPLTTAQ 2892
237 QY -----GT-----L--L-----A-----TG-----243
2893 Db AIIAKNPFDPVTPKGTPTIKDPKEFVDLEWTRPEADGGSPITGYVVEKREKSPDWEKC 2952
244 QY -----S-----Y-----245
2953 Db AEISDDITIAHVPDLIEGLKYEFRVAVNAKAGSPSDATETHVARPKNTPPKIDRNFMS 3012
246 QY -----D-----G-----F-----248
3013 Db DIKIKAGNVFDPVVTGEPPLSKDWTHEGNMIINTDRVKISNFDRTKIRILSATSDTG 3072
249 QY -----AR-----I-----W-----T-----253

3073 Db VYTLTARNINGTDRHNKVTILDAPSPVPBALRGDVSXNSIVLWRWPKDKDGGSEITHY 3132
254 QY -----K-D-----G-----NL-----A-----S--T-----261
3133 Db VVEKMDNEARWVPVGDCTDTEIRADNLNLDHDSYFRVRAVNKQGOQOPITTSQPI TAKD 3192
262 QY -L-----GO-----H-----K--G--PI-----F-----A-----271
3193 Db PYGHPDKPOQOATDQWKEHFVDLEWSTPKRDGGAPISSYIEKRPFQGWERAUVLGDN 3252
272 QY -----L-----K-----273
3253 Db CKAHVPELTNGEYEFVRVIAVNRGGSPDPSSTIICKPRFLAPFDFKSLNDITVHAG 3312
274 QY -----W--N--KK-----G--N--F--I--L--SA-----284
3313 Db NGLGWTLPTEASPRPLITWLYNGKEIGSNSRSGSLFQNELTPEIIVSSLSRSABGRYTLIL 3372
285 QY -G-----V-----D--K--T-----T-----II-----292
3373 Db KNEHGSFDSAHATVLDLRPSPPKPLDITKITRDGCHLTWNVPDDDDGGSPILHVIIEKMD 3432
293 QY -----W-DA-----H--T-----G-----AK-----301
3433 Db LGRSTWSDAGMSTHIVDVTRLVHRKEYLFRVKAVNAIGESDPLEAVNTIIAKNEFDEPD 3492
302 QY -----Q-----302
3493 Db APGLIITDWRDHDLDQWAVPKDGGAPISEVIIQKKEKGSFYWTNVRVPSNKNVTII 3552
303 QY -----O-----P-----P-----FH-----307
3553 Db PELTEQGEYEFVIAVNAQOSEPSPDMIMKRPYLPKIIITPLNEVRIRKGLIFHTD 3612
308 QY -----307
3613 Db IHFGEPAPEATLNSNPLLSNDRSTITSIGHHSVVHTVNCQSDSGIYHLLLRNSSGI 3672
3673 Db DEGSFELVLDLRPGPEGPMEEITANSVTISWPKPKDNGGSEISSYVIEKRLDTHGGG 3732
308 QY -----SA-----P-----A-----L--D-----V-----314
3733 Db WVPVNVYSKYNHAVVPRLLLEGTMVELRVMAENLQGRSDPLTSDQPVVAKSQYTVPGAP 3792
315 QY -----D-----W-Q--SN-----NT-----321
3793 Db GKPELTDSKXNHIITIKWKQPI SNGGSPITGYDIERRDVTNGRWIKINGQVPTAEYQDDR 3852
322 QY -----322
3853 Db VTSNHQYQVRISAVNAAGKGTSEPSAIFNARPLREKPRFYFDGLIGKIKVRAGEPVNL 3912
323 QY -----A-----SC-----325
3913 Db NIPISGAPPTTIEWKRGLKLEEGKRI SYETNSERTLFRIDDSNRDRDSKGYTVTAANERG 3972
326 QY -----S-----T--DM-----329
3973 Db KDTADIEVIVDKPSPGPGPLSYTETAPDHISLHWSPKDDGSDITGYIIEFTBFGVDD 4032
330 QY -----C-----330
4033 Db WKPVFGTCFNTNFTVKNLVEGKKYVFRIRAEANIYGASEALEGKPVLPKSPFPPGAPSOP 4092
331 QY -I-----H-----V-----C-----334
4093 Db TISATYPSANLEWPPDDCGGKPIITGYIVERRRERGERGWIKNYPTNTSYTVSNLRDG 4152
335 QY -----K--L-----G-----Q-----DR-----P-----341

Db 4153 ARYEFRLAVNEAGPHGSKSPDMPTAEHQYRPPPEPKPDRIITRNGVTLSWRPRTD 4212
QY 342 --IK--T--F-- 345
Db 4213 GKSRIKGYVEMRPMKDKWTDIPINSTVTVPFSLKEGEBYSFRVAENEVGRSDPS 4272
QY 346 --Q--G-- 347
Db 4273 KPSQPTIEEQPNKPCWELGKVRDIVCRAGDDPSIHVPYLLAPKPNAPFWYNDNMLDDNN 4332
QY 348 --H-- 348
Db 4333 RVHKLHTDDAASVVVVKSRKRDGQYRLQKDTSGEDTATINVRVLDSPPTLRADVF 4392
QY 349 --T--N-E-- 351
Db 4393 SGDSLTYWNPNDGSAIQNYIEKKEARSTWSKVSFCTVLFVRNLVLNKEYDF 4452
QY 352 -V--N--A--I-K--W--D--P-- 359
Db 4453 RVIAENKYGSDPANTSEPIARHPFDIPNTPGPHGIDSTEDSIATWTKPKHGGSPI 4512
QY 360 TG--NL--L--A--S-- 366
Db 4513 TGYIEKRLSDDKWTKAVHALCPDLSCKIPNLIEAIEYFRVAAVNAAGQSAYSDDL 4572
QY 367 --C--SD--DMT--LKI--W-- 376
Db 4573 IFCRRPPHAKTISDLSIPDMTVIAGDEFRITVPHYASPRPTASWSLNGLEVIPGERIKF 4632
QY 377 --SM--KOD--N--CVH--D--L--Q-- 388
Db 4633 DSNDYASMYNKSARDETSYITILTNKSGDTASC-HVTVDRLPLPGQPLNAVYDTP 4691
QY 389 --A--H-- 390
Db 4692 DTCTLAWKPLDDGGSPITNYVVEKLDNSGWSWKISSFVRNTHYDVMLEPHYKYNFRVR 4751
QY 391 --N--K--E--IYT-- 396
Db 4752 AENQYGLSDPLDIEPMVAKHOFTVPDEPGQPKVIDWDSGNVTLIWTRPLSGGSRIOQY 4811
QY 397 -IK--W--KTYQLYNLINGSEYEFRIKAKNAAGLSKSPSPSLRFLK 4871
Db 4812 QIEYRDILNDSSNAYDYIIKGTQYLYNLINGSEYEFRIKAKNAAGLSKSPSPSLRFLK 4871
QY 403 -G--PG--T--N--NP-- 410
Db 4872 KGKFTVPSPGAPQVTRVGKNTYDLKWEKPLRDGGSRIITGYIERRDIGGAVVVKCNDYN 4931
QY 411 --ANL--M--LA--SA--S-- 419
Db 4932 VLDTEYTNWNLIEMGDYFRFRVAVNSAGRSFSLCTMPIKCEVLGGKKPDWITRLQDKV 4991
QY 420 --F--D--S--TVRLW--DV--D--R-- 431
Db 4992 APFGKDYTLQCAASGRKPSPTAR-WLRNGKEIQMNGGRMTCDKDGVRFLHISNVQTDGDO 5050
QY 432 --QEP--G--I--C-- 434
Db 5051 DYTCEAMNSLGFVNTSGYLKIGSPIINRCPSBLKLPEDGNSKIKIPYSGDQPLTVILKKN 5110
QY 435 --I--H--T--LT-K--H-- 441
Db 5111 NEVISDSNDTHVKVNIFFDYVAIYANIVKSDGGPYQIEFTNESGATGEFYVHTGMP 5170
QY 442 --QEP--V--Y--S--V--AP-- 450
Db 5171 SAPTGMGILYNKNSCMLNWRPPYDGLKVSHTYTERKDVSSPHWITVSTCKDTAFN 5230
QY 451 -- 452
Db 5231 VQGLIENQBYIFRVMAVNMNGMPPLEGLNPIRAKODIDPPSPFPAVLKSLRSEETLCNPE 5290

QY 453 -----D-----GR-Y----- 456
Db 5291 WKPPSDGGAHQGYWIDKREVGSKHMGVNTATICAANOINCINLEGRQYEFRIQAQNV 5350
QY 457 --L--A--S-GSF--D-K--CV--H-- 469
Db 5351 AGLSTAIVRFTSSQDNRSSTGGFASIDCETATHANCIONHNAQFTCTIPGVKPTISWYKG 5410
QY 470 ----N--TQ--T-----GA----- 475
Db 5411 AREISNGAGITCTOKVTSPLPKYYVFGDEADDEVCRVAVNKKAGAKSTRATLAIMTAPKLV 5470
QY 476 -----L--VH-----S--Y-----R- 481
Db 5471 PPRFRDTAYFDKGVNVVIKIPFTGLPKPRIHVHVGLENIESGGHYTVVEKHERHAVLIIRD 5530
QY 482 G--TG-----G--I-----F-----B-----V--C 490
Db 5531 GSHLDGSPYRITAENELGSDTAIIQVOISDRDPPPRFPLIESIGTSTLSLWKPAPVWDGC 5590
QY 491 -----W--N--A-----A-----G-----D-- 496
Db 5591 SDITNYYVERRDDPLSSWIRVGNTRFTSMVSLTPGKEYDFRIFADNVYGRSDASDTST 5650
QY 497 ----K--V-----G-A-----S-AS- 503
Db 5651 LIKTESVKKKPIERKWEIDANGKRLGKADGPKVDYSDYVFDIYSKFPQVPQVEISQOSV 5710
QY 504 ----D--GS-----V--C-----VL-----DL--RK 514
Db 5711 YDRYDILEIGTGAGFVHRCRSTGNIFAAKFPVSHSVKOLIRRE 5759

RESULT 5

T30192

probable peptide synthetase - Aureobasidium pullulans

C.Species: Aureobasidium pullulans

C.Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004

C.Accession: T30192

R:Peery, R.B.; Thorneswell, S.J.; Tobin, M.B.; Skatrud, P.L.

submitted to the EMBL Data Library, January 1997

A.Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aureobasidium

A.Reference number: Z20767

A.Accession: T30192

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-10797 <PEE>

A.Cross-references: UNIPROT:O94116; EMBL:U85909; NID:g4099310; PID:g4099313; PIDN:AAD0056

C.Genetics:

C.Introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2

C.Keywords: carrier protein

F:1618-1688/Domain: acyl carrier protein homology <ACP1>

F:3682-3752/Domain: acyl carrier protein homology <ACP2>

F:5615-5685/Domain: acyl carrier protein homology <ACP3>

F:7503-7573/Domain: acyl carrier protein homology <ACP4>

F:9683-9752/Domain: acyl carrier protein homology <ACP5>

Query Match 70.7%; Score 2593.7; DB 2; Length 10797;

Best Local Similarity 8.5%; Pred. No. 8.2e-52;

Matches 461; Conservative 40; Mismatches 8; Indels 4923; Gaps 386;

QY 1 M-----SI--S-----S-----D----- 6

Db 4302 MLNTQVIOQFGSGIKSGILYGMGYPTAAIHCTLOPGFVDLPAGTIGIPDITVSCFIV 4361

QY 7 -----E-V-----N-----FLV-----YR-- 14

Db 4362 KPESTKHSQLEILPIGELVIGGHQLADGYLNREBQTRAAF-VTHPKFGLYRTGD 4420

QY 15 -----Y-----L-Q-----E-----SG-----F-- 21

Db 4421 KARLHRNGTLECYGRISGQVKLRQORVELGEBIEHAASKAGGCHAVIASVISGLLVLPFI 4480

QY 22 --SH-----SA-----P-----T-P----- 28
Db 4481 GDPRVSSKDIKSAQCKQKPAYMIPSDIVLDDPPYLPSCVKDKRLETDYNSNTAQHVS 4540
QY 29 G-----IES-----H-----I--SQ----- 36
Db 4541 GSSDLSNAREITRIIESVLGVSIDHSTDLASAAGLDSLRAIQVASQLRRQCADLGALEL 4600
QY 37 -----SNIN-----G----- 41
Db 4601 LSVSNVLALDELVRAKADESNINDNDSEKWKQTVHELRSVERDFESKAFVSGIEDVLP 4660
QY 42 -----A-LV-----P-P-----A-----AL-----IS- 51
Db 4661 TPLQDAMLVETAKPQAYCNELRLTVSPKIPVERVROALFALAQHRTALRSGMPGVS 4720
QY 52 -----II-----Q-K-----G-----L-QY-----VEAR--V----- 64
Db 4721 CAVTQVIWKTIVTSQFAHVKSFVTGMSVTNRETLLRPLHFQYKCSGAEAILVAIHVALY 4780
QY 65 --S--I-----NE-----D-----D-G- 70
Db 4781 DQMSVEILEDLETLQONERTPERPSFGAVNKFNLRRSEDQTSHLDPWGEYLSVTPGR 4840
QY 71 -----TL-----F-----D- 74
Db 4841 LPNLSPKMPPPQLOSIOHTIEMDMETLROAAHSYSCSAHVFPQAAVAILLGYMGTEDT 4900
QY 75 -----GR-P-----LS-L--I-----DA-----VM----- 88
Db 4901 VFGTVSGRTLPIVEIESMVGPLSTLPTINTLESKRKFSDSLRLQEDNRKIMRHSMTS 4960
QY 89 -----P-----D-V-V-Q-T-----R-----QO- 97
Db 4961 LADIKKACGNPEAVPDSIFVQWETARPDPARAQTLNLVNEADHYLEFNLTLBLEPTQOQ 5020
QY 98 --A-YR-----D--KL--A-----Q-----Q- 106
Db 5021 VKTKATYQSSLLPLQHVKTLLQOLDALVKIIVARPETHMNEISDQLPISVLVANSPEQS 5080
QY 107 -----HA----- 108
Db 5081 FVYKAGLSIVENHALNMSGGLALVFAHDIRECTSRMESITYGELNTRANQANLYISQ 5140
QY 109 -----AA-----GE-----AA----- 112
Db 5141 AKRDELICVMEKSVSLYLSILAAVKAQCYLPLVPETPAARIRQILAEADVFKCLTDS 5200
QY 113 --A--A-----A-----A-A-----A-----T-- 118
Db 5201 MAPVIADVSRCHIMNVDTDCSAQSGCTGPOLDKPTDIAVAVFTSGTTGPKGVLTQEN 5260
QY 119 --N-----Q-Q-----Q-----G-----SA-----K- 125
Db 5261 ILSNLEVLSKIYPVPEGSRLQACNOAFDVSVPEIFFTWYTMCLCSASKDMVFRDFEKA 5320
QY 126 -N-----GE----- 128
Db 5321 INELEITHLSLTPVAALTDPAHPRVKFLVTAGEAVTHHVHAGWAKGLYQGYGPSETT 5380
QY 129 N--T-----A-----N-GE--FN-----GA----- 138
Db 5381 NICTVNSAVESDHVINNIAPFENTSAFVLTOGDDFQLVPLGGLGELCGFGQOQVFRGYN 5440
QY 139 -----H-----TT----- 141
Db 5441 MPELTSKINHPNYGRIVRSGDLGRLLPDGTLIQGRTDDQKIRGORIELGESCGL 5500
QY 142 -----AN-----N-----H----- 145
Db 5501 QPFSVQNCALIEVKTADKRLMAFIPSGYSKDSYSILOPKMLBIIKSIYAHILADNLP 5560

QY 146 -----T-DM-----M--E-- 150
Db 5561 AYVVPDALVPVSAIPOTSQKIDKRRLASDGSALTVEDLNAYSRGADDDTSELSATEQQ 5620
QY 151 -----VD-----G-D-V-----E-----I-- 157
Db 5621 LASALADTLQMSQTSIGRSTSPFALGLDSVAIRLATNLNRKEYGYSIDVSQLKRRTIAR 5680
QY 158 --P--S--NK-----AV-----VL----- 165
Db 5681 LAPLLGGESSKOTNEPVTADCEAAAVGYLHDESIVVSQLHEHGQTVSQVLPCTPQEAMLS 5740
QY 166 -R--G-----H-----E-----S-E--V-PI----- 174
Db 5741 ARDTSSSAYRNKTLFSLHGSVDKLCACWEVMLQRHDIILRTIFLSTEDSRFPFVQAVLSQ 5800
QY 175 -----C-----A-----W----- 177
Db 5801 WTLPMQECDDIPDQLSTLLDSAKAGGDSIVDHPMPWKIQVYRSESTVYLLDMHHALYDA 5860
QY 178 --N-----PVS----- 181
Db 5861 NAMSLLLYEVEQLYKQDQSLSAFVSPKPLNFMISTSVBEADALFRDQLREFVPKPKRTD 5920
QY 182 ----- 181
Db 5921 VKSGFGTITGRNLNYSPPKVVETFLSKHSTTMLSITQAMWMTLAASQSYSDVCGNVVSGR 5980
QY 182 -----DL----- 183
Db 5981 SVPVDGIESLVAFCFNTPVRVDLSKHSNGLVKALQRVNDISLPYQLTPLRRIQAQAG 6040
QY 184 -----LV-----S-----G-SG--D-----ST----- 192
Db 6041 TNGKRLFDLSVLQQDTTDLDSAIWRLEGESGMDMYTSGTGLPKAVCLSHRAVTSQSL 6100
QY 193 A-----R-----I--W-----N-- 197
Db 6101 AHDRTIPSRFLQFASPTFDVSVFEIIPPWYRGATLVSVERNLLGDLPGTITSNIDA 6160
QY 198 --L--S--EN-----S-----T-----SG----- 205
Db 6161 ABELTPSAASLVHRHENVPTLRALLTIGEMLNTQVIOQPGSIGSGILYGMYPTEAAI 6220
QY 206 -----PT----- 209
Db 6221 HCTLOPGFVGLPAGTIGIPLDTVSCFIVKPTSTKHAQLEILPIGEIGELVIGHQLA 6280
QY 210 -----V-----L-----R-H--C--I-----R----- 216
Db 6281 DGYLNREEQTRAAFTVTHPKFGLYRTGDKARLHRNGTLECYGRISGQVKLRQORVELGE 6340
QY 217 -E--GG-----Q-----D-V----- 222
Db 6341 IEHAASKAGGCHAVIASVISGLLVLCIGDPHRVSSKDIKSAQCKWLPAYMIPSDIVLD 6400
QY 223 --P-----SN-----KD-----V-----T----- 229
Db 6401 DFPYLPBGKVDKRLKETDYNSTAOHVSGSSDLSENAREITRIIESVLGVSIDHSTDLA 6460
QY 230 -----SL-----D--W-- 233
Db 6461 AGLDSLRAIQVASQLRRQCADLGALELLSVSNVLDELVRAKADESNINDNDSEKWKQ 6520
QY 234 -----N-----S----- 235
Db 6521 TVHELRSVERDFESKAFVSGIEDVLPCTPLQDAMLVETAKRPAQYCNELRLTVSPKIPV 6580
QY 236 E-----G-----TL-----LA-----TG----- 243
Db 6581 ERVROALFALQORHTALRSGFMPSPGVSHCAVYQVIWKTIVTSQFAHVKSFVTGMSVTNRE 6640
QY 244 -----S-----YD-----G----- 247

6641	TLRLPLHFQKCSGAEBAILVAIHIALYDOWSVEVILEDLLETLQNERTPERPSFGAVNK	6700
248	FA-R-----I-W---TKD---G---NL-----	258
6701	FFNLRRSEDQSHLDFWGEYLS-DVTFGRPLNLSPKMPPQPLQSIQHTIEMDMETLROA	6759
259	-----A-----S-TL-----G-----	263
6760	AHSYSCSAHVFFQAAAYAILLGYMGTEDTVFGTVFSGRTLPIVEIESMVGPLLSTLPTRI	6819
264	-----Q-----H-----K-G-P-----IFALKW-----	274
6820	NLESRKFSVDLSRLOEDNRKIMRHSMWTSADIKKACGFNPGEAVFDSIFV--WQETARP	6877
275	-----N-----K-----F-----	276
6878	DARAOITLLNVEAHDYLEFNLTLELEPTQOQVKTKATYQSSLLPLQHVKTLLQOLDALVK	6937
277	-----K-----G-----N-----F-----	280
6938	IVVARETHMNEISDQLPISVLSVANSEPOSFVVKAGLGSIVENHALNNSGGLAVFAHD	6997
281	-----I-----J-----IS-----AG-285	
6998	IREGTRMESLTYGELNTRANQLANYLISQAKRDELICVMEKSVLSYLSILAANKACG	7057
286	-----V-----D-----	287
7058	GYLPLVPETPAARIRQILAEADVKFCLTDSMAPVIADVSRCHIMVDDTDCSAQSGTGP	7117
288	-----K-----TT-----I-----	291
7118	QLDFKPTDIAYAVFTSGTCKPKGVLTQENILSNLEVLSKIYPVPEGRLLQACNQAFD	7177
292	-----I-----W-----D-AH-----	296
7178	VSVFEIFFTWYTGWCLCSASQDVFRDFEKAINELEITHLSLTPTVAALTDPAHPRVKF	7237
297	-----K-----Q-----	302
7238	LVTAGEAVTHVIGAWAGKGLYQGYGPSETTICTVNSAVESHVNNIGPAPENTSFAV	7297
303	-----Q-----P-----F-----H-----SA-----P-310	
7298	LTQDDDFQLVPLGGLGELCFGGQVFRGYQNMPELTESKIINHPNYGRIYVSGDLGRLLP	7357
311	-----AL-----D-----V-----D-----W-----	316
7358	DGTILIQRTDDQKIRGORIELGEISGCLLPFSPVQNCIAEVIKTADKERLMAFWIPSG	7417
317	-----Q-----S-----N-----	319
7418	YSKDSYSILOPKNLEBIIKSIYAHADNLPAYVPDVALPVSAIPQTSQGIKDKRLAS	7477
320	-----N-----T-----FA-----	323
7478	DGSALTVEDLNAYRGADDDTSELSATEQOLASALADTLQMSQTSIGRSTSPFALGLDS	7537
324	-----S-----C-----S-----T-----DM---C-330	
7598	HDESVSVQLHEHGQTVSVLPCTPLOEAMLSARDTSGSSAYRNKTLFSLHGSVDKLKACW	7657
331	-----I-----H-----V-----C-----K-----L-----G-----	337
7658	EVMQLQRHDLITFLSTEDSRFFVQAVLSQWTLPWQECDDIIPDQJLSTLDSKAGGDSI	7717
338	-----Q-----DR-----PI---K---343	

Db	7718	VDHSPPWKIQVYSESTVYLLDMHHALVDANAMSNLLYEVEQLYKDQSLSAPVSPKPF	7777
Qy	344	-----T-----F-Q-----G-----H-----	348
Db	7778	NFMISTVEEADALFRDQLREFVPKPKRTDVKSGGTITGRNLNYSPOWVETFLSKHSTT	7837
Qy	349	-----T-----N-----E-----V-----	352
Db	7838	MLSITQAMMKTLTAAQSQSYDVCNGNVSGRSVPVDGIESLVAPCFNTIPVRVDLSKHS	7897
Qy	353	N-----A-----I-----K-----W-----	357
Db	7898	NLGLVKALQRWNDSLPYQLTPURRIOAQAGTNGKRLFDSLVLVLLQDTTDLDSAIWRLEG	7957
Qy	358	-----D-----PT-----GNL-----L-----ASC-----SD	369
Db	7958	ESGVMDPPCIVELAPNTESYTLSHFNRSYLDDEVSNLHQACLSAFASCIRYPSSDVS	8017
Qy	370	-----DM-----TLK-----I-----WS-MK-Q-----	380
Db	8018	FIDFDADLVAGLVKPDTKHMQPVEAAKTRNSEKSGSGDESWSPLELQIRAAAYSAVSSAP	8077
Qy	381	-----D-----N-----C-----V	384
Db	8078	EDRIRRDITYKGLDSISAIQLANRLRKDGLLVQASDVWESPSCSELASAVQSRQTPV	8137
Qy	385	-----H-----DIQ-----A-----H-----N-----	391
Db	8138	LDERGDFEGDKHYGAALQSHRIATEKVASVRPCTPLQSGMLSEYTHSDGHOYFNHTF	8197
Qy	392	-----K-----E-----I-----YT-----I-----	397
Db	8198	YAIEADIDSKLOSAMSKVLEQHELLRTGVTGDDHHPVMLTYTFDFVIDLEIQASSR	8257
Qy	398	-----K-----WS-----	400
Db	8258	EGSVYEYSEBOKASESVKNNLHLPWPWRWSLLGVGEGRCLOFSAHHAIFDAESLIMTDLQ	8317
Qy	401	-----PT-----G-----P-----	404
Db	8318	SALSNGVYPTRLTIDGALGHILNSQADVESQRTFWSQKLSGAPVTRFPNMVTRISDTE	8377
Qy	405	-----G-----	405
Db	8378	AANVELVNLKRSKLEARCQELGVMQSVQAAWARLLSAYTGESQVTFGWLSGRTS	8437
Qy	406	-----TN-----N-----P-NA-N-----	412
Db	8438	TADAAPFCITTLPVSTNTAVDDSQFLKDLMSYNATIQKHQFTPLTNIRNYAESTSEALFD	8497
Qy	413	L-----M-----LA-S-----A-S-----	419
Db	8498	SLFVYQRPMDVVDSWKIIREKASVELAVSVEALSEDGLGLRLTIDPAQVPEHQK	8557
Qy	420	-----F-D-----S-----T-----V-----R-----	425
Db	8558	IMLQQMEVMIALKKFEADINTSVMSIIPPKOPIATDFKYLHEMTEASVKSYS	8617
Qy	426	L-----W-----D-----VDRG-----	432
Db	8618	FVDALEDGQTSRRHWYRQLDEEANKIAHLLIDRGVKGPDIIATSPDKCEASFAFYGIL	8677
Qy	433	-----I-----	433
Db	8678	KAGCAFCADPTAPARKAFILEDSDNARVLITSDSIRSELRELTCQDIIIDLINFEKNEL	8737
Qy	434	-----C-I-----	435
Db	8738	STSSVPVSGLAPSSVSVLYTSGTTGPKGEITHDNAYQLVMSFKRLFKGRWTD	8797
Qy	436	-----H-----T-----LTK-----	440
Db	8798	OFASYHFDVSVLEQFWTIWGMRLVCAPRDLILEIAGFDLTMOITPHLDTLPSLGRLLDP	8857

```
QY 441 -----H-----Q-----E-P---444
Db 8858 ALVPSLHKGVFTTGESLKQDQNTWGDVGCLFNFGPTTECTIGTVTFVPCVKPKGPSNI 8917
QY 445 -----V-Y-----S-V-----AF---450
Db 8918 GWQFDNVGVVLAPGQTQPLVRLGAI GELCISGKLVGKGYLNPRLTADCFPYLDAGGERV 8977
QY 451 -----S-----P---452
Db 8978 YRTGDLVRLFHDSIDFLGRKQNVKLRGORLEIDEIAVIRKQDIOQTVICVAKHPKQ 9037
QY 453 --D-----G-----454
Db 9038 DKDQLIAFTIGNESRKQKPELCPABSTRHLIQTAARACEERLPGVMVTHFLPIQRIPL 9097
QY 455 -----R-----Y-----LA-----458
Db 9098 SVNNKVEKLLRQLYADLPVTIQTATQADSQSLSDGEGKVAQALAEKLLKIDNDLTP 9157
QY 459 -----S-G-----S-F-----D 463
Db 9158 SSNIFSLGLSSIAIQFSKKLKASGFTTVQVATVLRKNTPTISRLTKALATSTGRSGEIA 9217
QY 464 -----K-----C-----465
Db 9218 AKQVISACQRHMGTVTRVLRCKADIEIAIAPCTPLQOGLIISRLASESSLYFNPKFNA 9277
QY 466 -----V-----H-----I- 468
Db 9278 QGVDLQKLEGAFNQALERTQILRTFFIETDDGVVQAVRKTGHLPMWTLVYDLASVDGVF 9337
QY 469 -----W---NT-----471
Db 9338 AKRKQKRSYNTSHLTVPFEIVIVRSGETFVSVLDLHLYDGNSPDILMNNVSKLYNSQ 9397
QY 472 -----Q-----T-----473
Db 9398 EADFGKPFVDCLAFGLPRLNVQAKQFWLDLHPDKVASMPPLIDNPASHDVLCTASLDIL 9457
QY 474 -----G-----ALV-----HSYR---GT-----G-----I 486
Db 9458 NQADELRRLSGVTQALVQATWATLTKH-YQGAIGTVVSGRSIDFGVENVIGPLFNTI 9516
QY 487 -F-----E---V-C-----W-----491
Db 9517 PFYLRCEPGDTWQTLVQRCHDFNTTALPYQHTPLRDIVKWNCKNGHGRPLFDALFVYQGT 9576
QY 492 -----N-----AAG-----D---496
Db 9577 DNSDTHSILKPLEDDSFADYPLSPFEAEAGGNLKIISVAAKASICNETKARELIDEPH 9636
QY 497 -----K-----VGAS-----A---SD---GS-----506
Db 9637 QAFPLANKSPENVGASIGHTFERLARTEDGTRKIQARDTSDFWGSEASVIRSEIATL 9696
QY 507 --V-----CV-LD-----L--R-K 514
Db 9697 AGVQEAIDERTSIFEVLGDSVDVAKLSRLK 9728
```

RESULT 6

I38344
titin, cardiac muscle [validated] - human
N/Alternate names: connectin
N/Contains: serine/threonine-specific protein kinase (EC 2.7.1.1-)
C/Species: Homo sapiens (man)
C/Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004
C/Accession: I38344; I38345; S20898; S20899; S63665; S37393
R/Labeit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A/Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.

A/Reference number: A57430; MUID:96026330; PMID:7569978
A/Accession: I38344
A/Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBI
A/Molecule type: mRNA
A/Residues: 1-26926 <LAB1>
A/Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:g1017424; PID:g1017425
R/Musco, G.; Tziatios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A/Title: Dissecting titin into its structural motifs: identification of an alpha-helix m
A/Reference number: I38345; MUID:95119041; PMID:7819249
A/Accession: I38345
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1977-2014 <MUS>
A/Cross-references: EMBL:X83270; NID:g602579; PIDN:CAA58243.1; PID:g602580
A/Note: conformation and properties are reported for a synthetic peptide corresponding t
R/Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A/Title: Towards a molecular understanding of titin.
A/Reference number: S20897; MUID:92258380; PMID:1582406
A/Accession: S20898
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 13597-14200, 'I', 14202-14696 <LAB2>
A/Cross-references: EMBL:X64698; NID:g37192; PIDN:CRA45939.1; PID:g37193
A/Accession: S20897
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: mRNA
A/Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A/Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
A/Accession: S20899
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: mRNA
A/Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2
A/Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
R/Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.
J. Mol. Biol. 256, 556-563, 1996
A/Title: Genomic organization of M line titin and its tissue-specific expression in two
A/Reference number: S63665; MUID:96177761; PMID:8604138
A/Accession: S37393
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 26729-26825 <COL>
A/Cross-references: EMBL:X92412; NID:g1236761
R/Gautel, M.; Leonard, K.; Labeit, S.
EMBO J. 12, 3827-3834, 1993
A/Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiat
A/Reference number: S37393; MUID:94008990; PMID:8404852
A/Accession: S37393
A/Molecule type: mRNA
A/Residues: 26831-26926 <CAN>
R/Improta, S.; Folitou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A/Reference number: A66736; PDB:1TIT
A/Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R/Pfuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A/Reference number: A66201; PDB:1NCT
A/Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C/Genetics:
A/Gene: GDB:TTN
A/Cross-references: GDB:127867; OMIM:188840
A/Map position: 2q31-2q32
C/Function:
A/Description: structural protein forming filaments in striated muscle
C/Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; prot
C/Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein
F:24752-25008/Domain: protein kinase homology <KIN>
F:94,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4068,4
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
tatus predicted
F:16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18

.21900,21935,22295,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248
F;26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 70.7%; Score 2592.2; DB 1; Length 26926;

Best Local Similarity 9.0%; Pred. No. 5.8e-51; Indels 4368; Gaps 365;
Matches 437; Conservative 58; Mismatches 12;

QY	1	M	-----SISD-----EV--N-----F--L-----	11
Db	20024	MRFNTEITAENLTINLKESVTADAGRYEITAANSSTTKAFINIVLDRPPTGPPVIS	20083	
QY	12	-----V-----Y-----R-----Y-----L-----OE-----S-----G-----	20	
Db	20084	DITEESTVLKWPCKYDGGSQVTNYLLKRETSSTAVTEVSATVARTMMKWLTTGEY	20143	
QY	21	-----F-----F-----F-----F-----F-----F-----F-----F-----	25	
Db	20144	QPRIKAENRFGISDHDSACVTVKLPYTTGPPSTPWNTVNTRESITVGMHPEVSVNGGSA	20203	
QY	26	-----F-----F-----F-----F-----F-----F-----F-----F-----	33	
Db	20204	VVGYHLEMDRNSILWQKANKLVIRTHFKVTIISAGLIYEFVVAENAGVCKPSHPSE	20263	
QY	34	-----S-----S-----S-----S-----S-----S-----S-----S-----	33	
Db	20264	PVLADACEPPRNVRITDISKNSVLSWOOPAFDGGSKITGYIVERRDLPDGRWTKASPT	20323	
QY	34	-----IS-----Q-----S-----N-----I-----N-----	40	
Db	20324	NVTETQFTISGLTQNSQYEFVRVARNAGVSIENPSEVGPITCIDSYGGVIDLPLEYTE	20383	
QY	41	-----G-----G-----G-----G-----G-----G-----G-----G-----	44	
Db	20384	VVKYRAGTSVKLRAGISGRPAFTIEHWKDDKELOTNALVCVNTTDLASILIKDADRLNS	20443	
QY	45	-----P-----P-----P-----P-----P-----P-----P-----P-----	48	
Db	20444	GCVELKLRNMAASATIRVQILDKPPGPGPIEFKVTVAETKILLWRPPADGGAKITH	20503	
QY	49	LI-----S-----S-----S-----S-----S-----S-----S-----S-----	53	
Db	20504	YIVEKRETSRVVSMVSEHLEBECIITTKIIGNEYIFRVAWVKYIGIEPLESDSVAK	20563	
QY	54	-----S-----S-----S-----S-----S-----S-----S-----S-----	53	
Db	20564	NAPVTPGPGPIEVTKITKNSMTVWSRPIADGGSDISGFLEKRDKXSLGFWKVLKETI	20623	
QY	54	-----QK-----GL-----Y-----V-----E-----E-----E-----E-----	61	
Db	20624	RDRQKVTGLTENSQYRVCVNAAGQGFSEPFSEFYKAADPIDPPGPPAKIRIADSTK	20683	
QY	62	-----A-----A-----A-----A-----A-----A-----A-----A-----	63	
Db	20684	SSITLWSKPVYDGGSAVTVYVEIRQGBEEBWTVTSTGEVTEYVYVSNLKPQVNYVF	20743	
QY	64	-----VS-----I-----NE-----D-----D-----D-----D-----D-----D-----	73	
Db	20744	RUSAVNACQGFIEWNEPVQAKDILEAPEIDLVALRTSVIAKAGEDVQVLIPFKGRPP	20803	
QY	74	-----D-----G-----R-----R-----P-----P-----P-----P-----P-----P-----	79	
Db	20804	PVTWRKDEKLGSDARYSIENTDSSLLTIPOVTRNDTGKILTIENGVEPKSSTVS	20863	
QY	80	-----S-----S-----S-----S-----S-----S-----S-----S-----	81	
Db	20864	KVLDTPAACQKLVKHSRGVTLWDPPLIDGGSPINIVYIEKRDATKRTVSVVSHKGS	20923	
QY	82	-----S-----LID-----L-----L-----L-----L-----L-----L-----	88	
Db	20924	STSFKLIDUSEKTPPFVFRVLAENEIGIEGPCETTEPVKAEPAPIRDLMSKDKSTSVI	20983	
QY	89	-----PD-----PD-----PD-----PD-----PD-----PD-----PD-----PD-----	100	

Db	20984	LSWTRPDFDGGSVITEYVYVERKKGQETWSHAGISKTCTBIEVSQLEQSVLEFRVFAKNE	21043	
QY	101	-----D-----K-----L-----L-----L-----L-----L-----L-----	107	
Db	21044	KGLSDPTVIGPTVKEIITPEVDLSDIPCAQVTVRIGHNVHLELPYKGPSPISWLDK	21103	
QY	108	-----A-----A-----A-----A-----A-----A-----A-----A-----	110	
Db	21104	GLPLKSEFVRFSKTENKITLSIKNAKKEHGKGYTIVLONAVCRIAVPITVITLGPSPK	21163	
QY	111	-----A-----A-----A-----A-----A-----A-----A-----A-----	112	
Db	21164	KGPIRDEIKADSVILSWDPEDNGGEITCYISIEKRETSQTNWKMVCSSVARTTFKVPN	21223	
QY	113	-----A-----A-----A-----A-----A-----A-----A-----A-----	117	
Db	21224	LVKDAEYQPRVRAENRYGVSQLVSSIIVAKHQFRIIPGPGPKPVIVNTSDGMSLTWADP	21283	
QY	118	-----T-----T-----T-----T-----T-----T-----T-----T-----	123	
Db	21284	VYDGGSEVTGPHVEKKERNLSILQKNTSPISGREYRATGLVEGLDYQFRVVAENSAGLS	21343	
QY	124	-----A-----A-----A-----A-----A-----A-----A-----A-----	128	
Db	21344	SPSDPSKFTLAVSPVDPGCTPDYIDVTREITLKNPPLRDGGSKIYGVYIEKROGNERW	21403	
QY	129	-----N-----N-----N-----N-----N-----N-----N-----N-----	132	
Db	21404	VRCNFTDVSEQYTVTGLSPGDRYEFRIARNAVGTISPPSSSGIIMTRDENVPPIVEF	21463	
QY	133	-----GE-----GE-----GE-----GE-----GE-----GE-----GE-----GE-----	137	
Db	21464	GPEYFDGLIISGESIRIKALVQGRPVPRVTFKQGVIEKRMMEITNVLSSTSLFVRD	21523	
QY	138	A-----H-----TI-----A-----N-----N-----N-----N-----	144	
Db	21524	ATRDHRGVYTVAKNAGSAAKAEIKVKQDTPGVVGPPIRFTNITGKMTLWMDAPLNDG	21583	
QY	145	-----H-----H-----H-----H-----H-----H-----H-----H-----	153	
Db	21584	CAPITHTIIEKRETLRALWALIEDCEAQSYTAIKLINGNEYQFRVSAVKNFGVGRPLDS	21643	
QY	154	D-----V-----EI-----PSN-----PSN-----PSN-----PSN-----	163	
Db	21644	DPVVAQIQYTVDPDAPGIEPPSNITGNSITLTWARPESDGSSEIQYILERRKKSIRWK	21703	
QY	164	VL-----R-----R-----R-----R-----R-----R-----R-----R-----	171	
Db	21704	VISKRPISRETRFKVTGLTEGNEYEFHVAENAGVGPASGISRLIKCREPVNPPGPTVV	21763	
QY	172	-----V-----V-----V-----V-----V-----V-----V-----V-----	183	
Db	21764	KVTDSTKTVSLEWSPVDPGDMBIEIYIEMCKTDLGDMHKVNAEACVKTRYTVTDLOA	21823	
QY	184	-----L-----L-----L-----L-----L-----L-----L-----L-----	195	
Db	21824	GEEYKFRVSAINGAKGSDSCEVTGTIKAVDRLTAPELDIDANFKQTHVVRAGASIELFTA	21883	
QY	196	-----W-----W-----W-----W-----W-----W-----W-----W-----	202	
Db	21884	YQGRPTPAVWSKPSNLSLRADIIHTDSFTLTVENCNRNDAGKYTLTVENNSGSKSIT	21943	
QY	203	T-----S-----S-----S-----S-----S-----S-----S-----S-----	213	
Db	21944	FTVKVLDTPGPPGPIITFKDVTGRSATLWMDAPLDDGGARHIIHYVVEKREASRRSMOISE	22003	
QY	214	-----C-----C-----C-----C-----C-----C-----C-----C-----	222	
Db	22004	KCTROIKVNDLAEGVPPYFRVSAVNEVGVGPEPEIVATEQPAPRRLLDVDTSKS	22063	
QY	223	-----PSNKD-----PSNKD-----PSNKD-----PSNKD-----PSNKD-----	233	
Db	22064	SAVLAWLKP-----DHDGGSRIITCYLLEMRQKSDLWVEAGHTKOLTFTVERLVEKTEYFR	22120	

QY 234 ---N---SE---G-T---LL---A---T---G--- 243
Db 22121 VKAKNDAGYSEPREAFSSVIEKEPOIEPTADLTGITNQITCKAGSPFTIDVPSGRAP 22180
QY 244 ---SY---D-G-F---A--- 249
Db 22181 KVTWKEEMRLKETDRVSITTKDRTLLIVKDSMRGDSGRYFLTLENTAGVKTFSTVVV 22240
QY 250 ---R---I---W---TKDG---NL---ASTLG--- 263
Db 22241 IGRPGVTGPVSVSSAESCVLWSGPEKDGTEITNIVKRES--GTTAWQLVNSV 22298
QY 264 ---Q---H---K---G---PI---F--- 270
Db 22299 KRTQIKVTHLTKYMEYSFRVSSNRFVSKPLESAPIIAEHFPVPPSAPTRPEVHVHSAN 22358
QY 271 A---L---K---W---NK--- 276
Db 22359 AMSIRWEEPHDGGSKIIGYWVEKERNILWKENKVPCLCNCYKVTGLVEGLEVQFRT 22418
QY 277 --- 276
Db 22419 YALNAAGVSKASESRPINAQNPVDAPGRPEVTDVTRSTVSLIWSAPAYDGGSKVGYII 22478
QY 277 ---K---G---NF-I-LS---AGV--- 286
Db 22479 ERKPVSEVGDRWLKCNYYITVSDNFTVTALSEGDTYFRVLAKNAAGVISKGSSESTGPV 22538
QY 287 --- 288
Db 22539 TCDEYAPPKAELDARLHGLDVTIRAGSDVLDAAVGKPEPKIIWTKGDKELDLCEKVS 22598
QY 289 ---T---T-II--- 292
Db 22599 LOYTGKATAVIFCDRSDSKYTLATKNASGTKAVSMVKVLDSPGCGKLTVSRTQ 22658
QY 293 ---W---D-A-H---T---GE--- 299
Db 22659 KCTLAWSLPDGGAEITHIYVERRETSRLNWVIVEGECPTLSYVTRLIKNEYIFRVR 22718
QY 300 ---A---K---Q--- 302
Db 22719 AVNKYGPVPESEPIVARNSFTIPSPGPIEVEGTGKEHIIOWTKPESDGGNEISNVL 22778
QY 303 ---T-II---QF---P---F--- 306
Db 22779 VDKREESLRWTRVKNKYVYVYDTRLKVTSLMEGCDYQFRVTAVNAAGNSEPSSRNFISC 22838
QY 307 ---HS-A--PALD---V---D-W-Q--SN- 319
Db 22839 REPSYTPGPPSAPRVDTTKHSISLAWTKPMYDGGTDIVGVLEMQEKDQDQYRVHTNA 22898
QY 320 ---NT---F--- 322
Db 22899 TIRNTEFTVPLKMGQKYSFRVAANVKNMSEYSESTAEIPEVERIEIPDELEADLKK 22958
QY 323 ---AS--- 324
Db 22959 VTIRAGASLRMLSVSGRPPPVITWSKQIGDLASRAIIDTTESYSLIIVDKVNRVYDAGKY 23018
QY 325 ---C---S---T-D--- 328
Db 23019 TIEAENSGKKSATVLVKVYDTPGCPCKVKEVSRDSVTITWEIPTIDGAPINNYIVE 23078
QY 329 ---M---C---I---H--- 332
Db 23079 KREAAWAFKVTTKCKSKTLRIISGLVEGTHYFRVLPENIYIGIBECPTSDAVLNVSEVP 23138
QY 333 ---V---CK--- 335
Db 23139 LVPKLEVDVDTKSTVTLAWEKPLYDGGSLRTGYVLEACKAGTERWVKVVTLPKTVLEHT 23198

QY 336 ---L---G---QD-R-P-I---KT--- 344
Db 23199 VTSLENGOYLFRIBRAQNEKGVSEPREFTVAVTVQDLRVLPITDILSTMPQKTIHVPAGRP 23258
QY 345 ---HT--- 350
Db 23259 VELVIPIAGRPAPPAASWFFAGSKLRESERVVTHTKVAKLTIRETTIRDTGTYTLEKN 23318
QY 351 ---EVNA---IKW---D--- 358
Db 23319 VTGTTSETIKVILDKPGPTGPIKIDEIDATISITISWEPPELDGAPLSGVVVEQORDAH 23378
QY 359 ---P---T-GN---LLAS---C-S--- 368
Db 23379 RGMPLVSESVTRSTFKFTRLTEGNEYVFRVAATNRFGIGSYLQSEVIECRSSIRIPGPP 23438
QY 369 ---D-D-MTL---K-I---W--- 376
Db 23439 ETLQIFDVSRDGMTLTWYPPEDDGGSQVTGYIVERKEVRADRWVRVNVKVPVMTYRSTG 23498
QY 377 ---S---M---K-Q---DN--- 382
Db 23499 LTEGLEYHRVTAINARGSKPSRSPKPIVAMDPIAPGCKPQNPRTDTTRTSVLSAWS 23558
QY 383 ---C---V---H--- 385
Db 23559 PEDEGSKVTGYLIEWQVDOHEWTKCNTPTPKIREYTLTHLPOGAEYFRVLACNAGP 23618
QY 386 ---D---L--- 388
Db 23619 GEPAEVPGVKTEMLEYPDYELDERYQEGIFVROGGVIRLTIPIKGPFPICKWTKQ 23678
QY 389 ---A---H---NKE---IYTIK--- 398
Db 23679 DISKRAMIATSETHELVIKEADRGDSGTYDLVLENKCKKAVY-IKRVIGSPNSPEP 23737
QY 399 ---WSP---TG---P---GT--- 406
Db 23738 LEYDDIQVRSVRSVRPPADGGADILGYILERREVPKAAWYTTIDSRVGTSLVVKGLKE 23797
QY 407 N---N---P-N--AN---L--- 413
Db 23798 NVEYHFRVSAENQFISGLPKSEEPVTPKTLNPPPEPSPNPPEVLTVKSSVLSWSRPK 23857
QY 414 ---M---L---A--- 416
Db 23858 DDGGSRTVGYIERKETSTDKVVRHNKTIQITTYTMTVGLVDAEYQFRIIAQNVDVGLSE 23917
QY 417 -S-AS---F---DS-T---V-R--- 425
Db 23918 TSPASEPVVCKDPFDKPSQPGLEILSISKDSVTLQWEKPECDGGKEILGYWVEYRQSG 23977
QY 426 -LW--- 427
Db 23978 SAWKSNKERIKDKOPTIGGLLEATEYFRVFAENETGLSRPRRTAMSITKLTSGEAPG 24037
QY 428 ---DV---D---R-G---I--- 433
Db 24038 IRKEMKDVTKLGEAAQLSCQIVGRPLPDIKWVFGKELIQSRKYKMSDGRTHLTVM 24097
QY 434 ---CI---H---TL--- 438
Db 24098 EEQDEGVVTCIATNEVGEVETSSKLLQATQFHPGYPLKEKYKYGAVGSTLRLHMYIG 24157
QY 439 ---T---K-H---O--- 442
Db 24158 RVPAMTWFGQKLLONSENITIENTEHTLVMKVRQKTHAGKYKVLQSNVFGIVDAI 24217
QY 443 ---E---P---V---Y--- 446
Db 24218 LDVEIQDKPKPTGPIVIEALLKNASAVISWKPDDGGSWITNYVVEKCEAKEGAWLV 24277
QY 447 ---SV---A---F---SP--- 452

Db 24278 SSAISVTTCTIVNLNENAGYFVRVSQNTFGISDPLEVSSVWIKPFPEKPGAPGKPTIT 24337
QY 453 -----DG---R---YL-----A-S-----G-- 460
Db 24338 AVTKDSVAVKPPASDGGAKIRNYILEKEKKQNKWISVTTBEIRETVFSVKNLEGLE 24397
QY 461 -SF-----D-----K-----C-V--H 467
Db 24398 YEFVRKCNLGGSEWSEIPEITPKSDVPIQAPHPKEELRNLRVYQSNATLVCKVTGH 24457
QY 468 -I--W-----NTO-----TG 474
Db 24458 PKPIVWYRQGEIADGLKYRQEPKGGYHQHIIASVTDGDDATVYQVRATN-OGGSVSG 24516
QY 475 -----A-----L-----VHSYRG-----T--G-- 485
Db 24517 TASLEVEVPAKIHLPKTLGEGWGHVHALRGEVVISIKIPFSGKPDVITWQGGDLIDNNGH 24576
QY 486 -I-----F--E-----VC----- 490
Db 24577 YQIVTRSETSLVFPNGVERKADGFFVYVCAKNRFGIDQKTVELDVADVPDPGRGVKVSDA 24636
QY 491 -----W-----N-----A----- 493
Db 24637 SRDSVNLWTEPASDGGSKITNYIVEKCATBRLRVGQARETRYVINLFGKTSYQPR 24696
QY 494 --A--G-----DK-----V----- 498
Db 24697 VIAENKFGLSKPSPEPIITKEDKTRAMNYDEVDRETVSWTKASHSTKELYEKYMI 24756
QY 499 -----G-----AS-----A-----SD----- 504
Db 24757 AEDLGRGEFIVHRCVETSSKTYMAKFKVKGTQDVLVKKEISILNIAHRNHLHES 24816
QY 505 -----G-----S-----VC-----V--LDLR 513
Db 24817 FESMBELVMIFEFISGLDIPERINTSAFELNREIVSYHVQVCEALQFLHSHNIGHFDIR 24876
QY 514 -----K 514
Db 24877 PENIYYQTRRSSTIK 24891

RESULT 7
T15789
hypothetical protein C41A3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15789
R:Bentley, D.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid C41A3.
A:Reference number: Z18404
A:Accession: T15789
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-7829 <BEN>
A:Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA831
C:Genetics:
A:Gene: C41A3.1
A:Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
P:5547-6616/Domain: acyl carrier protein homology <ACP>
F:2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 70.6%; Score 2590.9; DB 2; Length 7829;
Best Local Similarity 8.5%; Pred. No. 5e-52;
Matches 447; Conservative 52; Mismatches 11; Indels 4771; Gaps 356;
QY 1 MS-----I-----S-----S-D--E-----VN----- 9

Db 317 MSPSKSMGHGEGAAGLIGVLQSLYSWQHGIIPNQLHLELPSDLGEDKSMGFVNEEMELN 376
QY 10 -----FL----- 11
Db 377 RVAISSVGGGTNACALIEKPEKPSLVQKESYAESNVLFSLSAKSHESKLQIBEYTOPMA 436
QY 12 -----V--Y-----R-----Y-----LOSS-----G 20
Db 437 QSDSAMEDILYTVNERKTYDFRAAVFGKDNBEIARKLQGDYSLTLQBSTFEVEFEG 496
QY 21 -----FSSHS----- 24
Db 497 NEKMLLRMLYKNETP-HSTVDKYCKLAETCGPEPARTALFPFFKLTLPITYNVSRLLI 555
QY 25 --A--A-----F-----TF-----G-- 29
Db 556 SSWATFELLVQVNTLPNKLGRGLGQIFCLAVAKVITFESAVQLIKGVAAEANTDILGD 615
QY 30 -----IE-----S-----HIS----- 35
Db 616 IELKSSKIPTEIOHLKASTKKKILPIHISGELKETAKPNLMTFIVNGBEILELDPVRKVQK 675
QY 36 -----QSN----- 38
Db 676 LICQLFACGPDPAVKPRGRIVKTPYTSFLKQRFWEVQVTAMTNIVDEQTNSSLSDAIE 735
QY 39 -----IN--GA--L-----V--P-----PAAL-I 50
Db 736 STVTVTKQFLDIEEDDINLETGAVDSLTSIEMVEAFGTAVNQTFDFDLEAYPTILNI 795
QY 51 -----SI-----I--Q----- 54
Db 796 VDFLKTAVTVPTKATTSIHKKTSLSIDINVIACDYQFAGVEGKELMDTLTSLRT 855
QY 55 -----K-----GL--Q-----Y-V--EAE----- 63
Db 856 TGKSIDRKQCEGDAGLEVGLLKODISMFDSNFFAIAKDEAEFLDPQHRLLNAAYNAL 915
QY 64 -----VS-----INE-D--GT----- 71
Db 916 EKSGLTISPADLFLAISAHSEYRALAEKHINELDERLWGTVHSMVAGLAVLMGIRGR 975
QY 72 -----L-----F-- 73
Db 976 AMIVDTTCSVATALEMAVKSIREGKRFATVATSLIQSKWLYSLKTLDDHSTNSFSV 1035
QY 74 DG-----R-----P-----I-----E--SLSL 83
Db 1036 DSGGFCRSDGVGVILKTAEGDSAVIKISSAKSHHCGAVMTFVSVSISQLLEAGFSY 1095
QY 84 I-----D-----A-----V-----M----- 88
Db 1096 VEGHGTATGAGSAESMAYQKLGSELMSSVRAQFGHCEVASGLIQLMKVSSIKGKIIP 1155
QY 89 -----P-----D-----VVQT-R--Q-----Q 97
Db 1156 SIHVNLTPSEHIRNNENIRLPFAVEBKQIDRSATVSFGITGTGTVTTERVSVQLNDVNIQ 1215
QY 98 -----A-----Y-----R-----D 101
Db 1216 NCYLLPVSARKDGLKKACLSLIEMIDNSCESLYDISTTLQKQKTNFKWRTAVVGSASHAD 1275
QY 102 ---KLAQ----- 106
Db 1276 VVLKQKFLTSEHTSLTNWHISTSHSISGCGSTFFHNIPEFEDHYSMFCHRLRPFPSNT 1335
QY 107 ---H--A--A-----A-----AA--AA-----A-A-- 116
Db 1336 NSIYHLLAVVYALIRVILKHLTNSFVGVGFNSLIVLAAVDAAPSHVNLNLLHAFANDDV 1395
QY 117 ---A--T-----N-----Q-----QGSAN----- 126
Db 1396 RWMKRIARDVTISIENVKLLNNGEPITTARQAVEATIDQ-KVKNVRLPETTILSPSRA 1454

QY 127 -----GE-----N-----T- 130
Db 1455 YEPASOLETIQYKUIGEKFTQGGQVDFAGIFGTPIKILIDLPEYFNRKSFWLPIPDV 1514
QY 131 -AN--GE-----N-----GA-----H----- 139
Db 1515 PSNEKEKPLIPKSYBFLKSKQWQHVNVDKIVLPGATIRLVHQLNGKPTVELSN 1574
QY 140 -----T-----I-----A-----N-----N-----HT 146
Db 1575 IDFLNKITSEAPSVVKIEBQDGLKLVFGETDAIGFKLTELOFNPNIPNRLNAEVHHT 1634
QY 147 D-----MM-----E--VD-----G-----DV-----E- 156
Db 1635 DNIYERFANSHLYRNEFQWDSKYTMKGGEVRFVSMKDLIDLIDGTLOAIVGCYFFEN 1694
QY 157 -----IP-----S--N-----K--AV-----VL----- 165
Db 1695 TNDNSFVPTIDQLSILNGDISQQLHAVLVKYDSSGNFINGDATVYDALGNILHISNV 1754
QY 166 -----R--G-----H-----ESE-----V----- 172
Db 1755 TFXKLGQASPLTSKTVDSKITKVENEDQKRASKMHLHWFE-ENFGWTDIDNTGF 1813
QY 173 F-----I-----C-----W----- 175
Db 1814 FDLGLTSIQAVKLUNAISKYNPASSTCVDPYPSIDLLSGYLSTLNDPQVTTSTGEDDI 1873
QY 176 -----A-----W----- 177
Db 1874 QKDLTEDHKTRLAENPIGWMAAACRLPGGVSPSELWELLKIGKKNASSRIPATRVPTRN 1933
QY 178 -----NPV-----S-----D-----LL--V-----SG- 187
Db 1934 TLISGKYGNFVGGNFITQDVTQFDPFPFKISKSEALIDPQORLLLECVQCLENSGV 1993
QY 188 -----S-----G-----D-----ST-----A--R--I--W----- 196
Db 1994 IETSNVGVFVGLMEKYQDMMESSLAMLGSMAAVIAGRWNYIFGCGPSVTIDTACSS 2053
QY 197 -----NL----- 198
Db 2054 SLVALEMAINALLDNCRSKVIIVAGVNLILNEKQGLRTNGKMSLQHGMSLSPDSRAGYG 2113
QY 199 -S-----E--N-----ST--S-----G-----P-----T-----Q----- 208
Db 2114 RSDGCVVLMLELAKPNFHYMSTIOSVNVNNGRSVSLTAPNGVAHKMLLTSVINQPSLA 2173
QY 209 -----L-----VLR-----H-----C----- 214
Db 2174 IDYEAHGCTPLGDPFNTLSILQNIIGSVKASLHGCEASAGTCGLLKLFLMLTYQ 2233
QY 215 -----IR-----E-----B-----G-----G----- 219
Db 2234 YVPTLHFVNLKDNAGSIRLPIIGEDSELVAGISFVGSGTNAIAAFNDNNKLEPY 2293
QY 220 -----Q-----DVP----- 224
Db 2294 IPIHKYILPISAKNQISLDNLEKQILSVIPLTDVPICNIASALANNRSHFTIRNALIVS 2353
QY 225 N-----KD----- 228
Db 2354 NSGIVNSKMEGKPHRVAKDORYHVKLCDSLSDASLLQYOVINETYTVASLKNPQSPAMKF 2413
QY 229 -----TSL-----D-----W--N--S--E--GTL-----A----- 241
Db 2414 AIKFLTSLEYEIVASDGEELLAVLLANGSLKWNFNKMTIELPIGSLTTEFADHDLN 2473
QY 242 -----TG-----S-----Y----- 245
Db 2474 STTSSSIKSYQTQPSHNLDSPMELMKLIMKLYITGYDVVDWATVYSPVEQFIALPNYQFN 2533

QY 246 -----D--G-----F----- 248
Db 2534 KQTLMEERLEIVDHYLIGTIDEESDITLLKNQISELRHPQFFKCKPLDVGTMSEIAIE 2593
QY 249 A--R-----I-----W-----T----- 253
Db 2594 ALKIRNEIPFSIQNLKTELITLTKPAWLETNVTNRNEDDEGNVSAVIDGQRLFSLNASSV 2653
QY 254 -----K----- 254
Db 2654 EIQNIIEVPAVEQIPDKVVYLKECPNAVIRHRNMVYVDSRABQSPFRTANIVLEIIGF 2713
QY 255 -----DG-----N----- 257
Db 2714 APTPSDMFIEILGLPSVHYMVQVDDGALWQFOMISQDKRVLNIIYVLKDAKLEIPTIR 2773
QY 258 -----LA-----ST-----LG----- 263
Db 2774 MHKSTLLSSQEBASIVAAKTLQMAVRHKVCLAVGDVIESGLDIDESQLSTGFSLEGIDSL 2833
QY 264 ----- 263
Db 2834 ATVDLLNRLNKYFPEIELTSDLPDNPISIIDLSIMEQLLNEKGITSEPSEPNTPKTSR 2893
QY 264 -----Q-----H----- 265
Db 2894 GRKLSIPAVRAQVLAQIEFVENYNVKQEVQAEAPSSSECSNHLSESDATVDRTEIRK 2953
QY 266 -----KG-----P-----IF----- 270
Db 2954 VSLAVFDLATETLSAEDLQSKGTGELGMDLSIVDPVNRNDKYFPDDEITASDIDPYPT 3013
QY 271 -----A----- 271
Db 3014 VDELSDHIVRKSSSVPPAAASEIMKTMNGISTSDAETHKLENLSQSFMLLENQNSINP 3073
QY 272 -LK--W--N--K--K--GNFI--LSA-----G-----VD--K-----T-T- 290
Db 3074 TLMWNSQTIKLVKPSDGNFLFELNANGQOEKEIQKHFTGPNNIIIDLKGFHEGSTETL 3133
QY 291 -----I-----I--W----- 293
Db 3134 YMSLLNLVKSISKLEIQCRFGVSOEFLGNSISRAFMKTVAEKNPLISPAMYQNVQVS 3193
QY 294 -----D-----A-----H 296
Db 3194 FVDSDSPITGNWLITGSLGIGLEIGKFIANNGAENVILISRROPTAKALRDILSTELTH 3253
QY 297 TG----- 298
Db 3254 IGLAKTIVLKLINISAKLIVFQSPKLSFSTKKYISDLPSKKVTFPYNLLQSKCFSKHF 3313
QY 299 -----E-----A-----K-----Q----- 302
Db 3314 HMLFEWLEVYLIKQWTVHTTAADINDKEKLIBELTKLVNIGITGIHSAAGVLKDSKIERQN 3373
QY 303 -----Q--P--P--FH----- 307
Db 3374 KESFNQVFTPKANGFHVLEIEIKHFNYKIENFIMSSFTAAACNEGQNLGYVSNAYLEYQ 3433
QY 308 -----S-----PA-----A----- 311
Db 3434 VQRRRQSGGCAIQNGWIDTGMATDENVRKFLANGLFGQHNKALKYLACILTKPE 3493
QY 312 L-----DVDW-----Q-----SN----- 319
Db 3494 LIWVANIDWVILKNRKLDPKDLINTGILPFEDFTGMBESEPPLSGNDPEKYSMNFVSVE 3553
QY 320 -----N--TF----- 322
Db 3554 DEBEVLELIKESVSSILMCSPTKLKNKNIMDMGLSKLIVEFLNFINSTFKISVNLSDA 3613
QY 323 -----A-----S--C-----ST 327

Db 3614 YNPTLEKLAHIFEQMTIIVDHPVNSVKSEIEIFKSTDFCFIFGINIFFDKNDKAKST 3673
QY 328 -----D----- 328
Db 3674 AVKLENGEOLPTAGKYAVSVGKSRDVVSKIKKAPQOIKLCQSSSKVCLMLTGQGSQ 3733
QY 329 -----M-----C-----H---V--- 333
Db 3734 YPMGRQLVENEIEFRITLQSLCKCDEYLQGVSLWEILFNTDHYKLLQLTKHMQPIMF 3793
QY 334 C-----K-----LG 337
Db 3794 CFGYATAQLWSLGIYPDYLLGHVSGELVAGVLGIMSIEDGLRLIVERGKAMENIAGLG 3853
QY 338 -----O---D---R-----P---I---KT---F---QG----- 347
Db 3854 ALLAVOREIADEVLRKPKVSVATINSGKQVVFAGTKSVLDAALAFVKGQKQATYVNOQY 3913
QY 348 --HTN--E-----VN---A---I----- 355
Db 3914 PFHSNLIQETHVLSRQCLADIKFSAGRTPLVSNVTQIINTFSEAVIVKHTVSAVKFVD 3973
QY 356 -----K---W---D-----PT-----GNLL-A--- 365
Db 3974 CVETLOAKGVTVWIDAGSAAVLATFVKRIITOPTELSKHRIVQTCBESVDVNLVQACLE 4033
QY 366 -----S---C-----S----- 368
Db 4034 LEOSGLPISWTTLTGCGORNADERLVEFPWTHNDIINKDEPELLEGHQLNGKIVVAGAYQL 4093
QY 369 -----D-----DMT----- 372
Db 4094 FKIDQLVKLKAAGMELMKNVFLKWPYIEDNREYQIQWNSDMTIELIVNSVIVCSLEVE 4153
QY 373 --LK-----I----- 375
Db 4154 PQNSVLKLETISENEKPFVHDFYETLFRNGLOYDGFRRIESARRDKRCFQIKSSPF 4213
QY 376 -W-----S---MK-----Q-----D--- 381
Db 4214 ANPLIDSAMHSITASVVRPRPCYFLPVANGSVTMKDTNSFTLPLNLAQTVITSETDKFI 4273
QY 382 --N-----C-----VH-----D---L--- 387
Db 4274 QVNVALLAGDTPICEVNMNTIIVLKTPEVHTRIIPNSIETVETPPKSEIEIVGDISLPYN 4333
QY 388 Q-----AH-----N---KE---I-----Y----- 395
Db 4334 QISENENQHLKTNVTVKQKLHNSRLKQDHARVALLDSDARYWDPEYFGIRPSEAKFIDP 4393
QY 396 -----T-----I-----K-----WS----- 400
Db 4394 QORLLCSVAKLDSLILITSLTNTGVFGICSANEFSHIVYAYGYKDPRAENSGGTSNSA 4453
QY 401 -----P-----TG----- 403
Db 4454 LAGRIAHWLKLGPPVTLDTACSSSFYALSACDALRTGCEYAIIVGTNVNLVHMTVDV 4513
QY 404 -----P----- 406
Db 4514 LQNAKMTVDVDFCKAPVDANGYKRSBVCMSMLLTGSPNIDSVATITNYATGHNGTSSSLF 4573
QY 407 --N-----NP-----NA-----N--- 412
Db 4574 TPNGLSOLEVMORATPPEKILEIOTHCTGTGLGDPPIEINATSKLVSSACKIGSVKSNIG 4633
QY 413 -----LM----- 414
Db 4634 HTEGSGLSVLSLSSLSFRRSKRYVAQHLKCPNTSIKTNMKICRFGEADENNLSILNN 4693
QY 415 -----L---A-----SA-S-----F----- 420

Db 4694 FGFTGSCSVLKPKNNAISEHFVSSEVFPYILLSSHSAKSQKYQVQVLCFISNSAKSLH 4753
QY 421 D---S-----T-----VR----- 425
Db 4754 DIMMSLFQKKIHVHROPIIFNFKRKIAVTSLDGFEVVRDERLSEKLKHPCSVFLKEGVVHF 4813
QY 426 -----L-W-----D-VDR-----GI----- 433
Db 4814 DKDKSFQVDLPISIVFNNTLHWDLSYRDEIDRHSQOMSFKNI FYEKVLLETMPQONP 4873
QY 434 --CI-----H-----TL----- 438
Db 4874 ISKVCICGRDLIPKIEIDEVSSFANGIIIVFHPISNSIFEYLKLVKWSLISRNQNVFI 4933
QY 439 -----TKH-----Q---E---P---V-----Y----- 447
Db 4934 ICCFNGTSHTEWTGTLRSLSASEKMIYPKPVSDIKVDALKLEPNHEMFMFAIYFKRGY 4993
QY 448 -----V---AF-----S-----P-----DG----- 454
Db 4994 VBLRKVNPTFKPAPOVEKCLISGGTGIGSAIINELKPKSVIITRKNIASEDGKTFLS 5053
QY 455 --R-----Y---LASG-----SPD-----KC----- 465
Db 5054 SDITRLDISHKFNYPVFLA-GIVNNSLHENVKRDSLDEVSIKLOGAKNLMKCCDETSHF 5112
QY 466 -----VH----- 467
Db 5113 VFSSSTANVLGSGQSNYAFNSGLVTSFLETSTKSTIIHWGPWKDVGMLAQPEREIVK 5172
QY 468 -I---W---N---TQ---T---G---ALV---H----- 478
Db 5173 QIESNGMKLLPNQDAISVFYTFQMETHEQIIVFDGDFDIIVARQPHLQKLLSEVVEKTKV 5232
QY 479 -----S----- 479
Db 5233 KBEIKKSLNFEIPEIIVGITDISKLNIPFMDLIGIDSLCMENRLYSLNKNFDELTVS 5292
QY 480 -----Y-----R----- 481
Db 5293 EMPENATYKLOQTYVETLRKAKHSLSHDRVSSQVSNKEDDTRVAVIGWSABFSGSSNI 5352
QY 482 -----G---TG----- 484
Db 5353 HEYWNELMDGICSTGNKYLLKNPFGDNKFNLTDEDARVLDPOVRKFIQHAYLALENS 5412
QY 485 -----GIF-E-----V-----C----- 490
Db 5413 GYVKQKHELRCGVFAGABPSDYGRADDHDDAMRKLFVMMNNSYLSASYCLDLKGEAVS 5472
QY 491 -W-----NA-----AG-D-----KV-GA-----SAS----- 503
Db 5473 VYSACSTALVANAVKSIQSGMDYALVGAASIAEVSALSGDFDQKTMFKSGVCRP 5532
QY 504 --D-----GS-V-C-VL-----DL-----R----- 513
Db 5533 FDKDSGIVRSGVGVFLKRYSQLLDNDNVHFVVKDFAINNDGHSRASFMAPNAGQL 5592
QY 514 K 514
Db 5593 K 5593

RESULT 8

S25111

alpha-2-macroglobulin receptor precursor - mouse

N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C:Species: Mus musculus (house mouse)

C:Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 09-Jul-2004

C:Accession: S25111; S32554

R:Van Leuven, F.

submitted to the EMBL Data Library, July 1992

A:Reference number: S25111

A;Accession: S25111
A;Molecule type: mRNA
A;Residues: 1-4545 <VANI>
A;Cross-references: UNIPROT:Q61291; EMBL:X67469; NID:g49941; PIDN:CAA47817.1; PID:g49942
B;van Leuven, F.; Stas, L.; Raymakers, L.; Overbergh, L.; de Strooper, B.; Hilliker, C.;
Biochim. Biophys. Acta 1173, 71-74, 1993
A;Title: Molecular cloning and sequencing of the murine alpha-2-macroglobulin receptor c
A;Reference number: S32554; MUID:93250049; PMID:8485155
A;Accession: S32554
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-28,4416-4453 <VAN2>
A;Cross-references: EMBL:X67469
C;Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated
ciated protein (see PIR:JX0281).
C;Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
C;Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glycopro
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-3944/Domain: alpha-2-macroglobulin receptor #status predicted <WAT>
F;28-65/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;73-109/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;116-149/Domain: EGF homology <EG1>
F;155-189/Domain: EGF homology <EG2>
F;199-240/Domain: LDL receptor WYTD-containing repeat homology <YW01>
F;241-282/Domain: LDL receptor WYTD-containing repeat homology <YW02>
F;293-335/Domain: LDL receptor WYTD-containing repeat homology <YW03>
F;336-379/Domain: LDL receptor WYTD-containing repeat homology <YW04>
F;380-421/Domain: LDL receptor WYTD-containing repeat homology <YW05>
F;422-469/Domain: LDL receptor WYTD-containing repeat homology <YW06>
F;479-520/Domain: EGF homology <EG3>
F;572-614/Domain: LDL receptor WYTD-containing repeat homology <YW07>
F;615-660/Domain: LDL receptor WYTD-containing repeat homology <YW08>
F;661-711/Domain: LDL receptor WYTD-containing repeat homology <YW09>
F;712-753/Domain: LDL receptor WYTD-containing repeat homology <YW10>
F;754-800/Domain: LDL receptor WYTD-containing repeat homology <YW11>
F;808-843/Domain: EGF homology <EG4>
F;855-891/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;896-932/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;937-972/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F;977-1012/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F;1016-1052/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F;1063-1098/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F;1105-1141/Domain: LDL receptor ligand-binding repeat homology <LDL9>
F;1146-1183/Domain: LDL receptor ligand-binding repeat homology <LDLA>
F;1186-1222/Domain: EGF homology <EG5>
F;1228-1262/Domain: EGF homology <EG6>
F;1270-1309/Domain: LDL receptor WYTD-containing repeat homology <YW12>
F;1310-1356/Domain: LDL receptor WYTD-containing repeat homology <YW13>
F;1357-1399/Domain: LDL receptor WYTD-containing repeat homology <YW14>
F;1400-1446/Domain: LDL receptor WYTD-containing repeat homology <YW15>
F;1447-1489/Domain: LDL receptor WYTD-containing repeat homology <YW16>
F;1490-1532/Domain: LDL receptor WYTD-containing repeat homology <YW17>
F;1541-1579/Domain: EGF homology <EG7>
F;1584-1627/Domain: LDL receptor WYTD-containing repeat homology <YW18>
F;1628-1670/Domain: LDL receptor WYTD-containing repeat homology <YW19>
F;1671-1714/Domain: LDL receptor WYTD-containing repeat homology <YW20>
F;1715-1754/Domain: LDL receptor WYTD-containing repeat homology <YW21>
F;1755-1797/Domain: LDL receptor WYTD-containing repeat homology <YW22>
F;1798-1847/Domain: LDL receptor WYTD-containing repeat homology <YW23>
F;1851-1887/Domain: EGF homology <EG8>
F;1935-1977/Domain: LDL receptor WYTD-containing repeat homology <YW24>
F;1978-2020/Domain: LDL receptor WYTD-containing repeat homology <YW25>
F;2021-2064/Domain: LDL receptor WYTD-containing repeat homology <YW26>
F;2065-2106/Domain: LDL receptor WYTD-containing repeat homology <YW27>
F;2107-2152/Domain: LDL receptor WYTD-containing repeat homology <YW28>
F;2160-2195/Domain: EGF homology <EG9>
F;2200-2242/Domain: LDL receptor WYTD-containing repeat homology <YW29>
F;2254-2295/Domain: LDL receptor WYTD-containing repeat homology <YW30>
F;2345-2389/Domain: LDL receptor WYTD-containing repeat homology <YW31>
F;2390-2430/Domain: LDL receptor WYTD-containing repeat homology <YW32>
F;2431-2474/Domain: LDL receptor WYTD-containing repeat homology <YW33>
F;2483-2518/Domain: EGF homology <EG10>

F;2525-2562/Domain: LDL receptor ligand-binding repeat homology <LDLB>
F;2567-2601/Domain: LDL receptor ligand-binding repeat homology <LDLC>
F;2606-2640/Domain: LDL receptor ligand-binding repeat homology <LDLD>
F;2653-2689/Domain: LDL receptor ligand-binding repeat homology <LDLE>
F;2697-2731/Domain: LDL receptor ligand-binding repeat homology <LDLF>
F;2735-2770/Domain: LDL receptor ligand-binding repeat homology <LDLG>
F;2775-2813/Domain: LDL receptor ligand-binding repeat homology <LDLH>
F;2819-2854/Domain: LDL receptor ligand-binding repeat homology <LDLI>
F;2859-2898/Domain: LDL receptor ligand-binding repeat homology <LDLJ>
F;2905-2940/Domain: LDL receptor ligand-binding repeat homology <LDLK>
F;2945-2981/Domain: EGF homology <EG11>
F;2987-3022/Domain: EGF homology <EG12>
F;3030-3069/Domain: LDL receptor WYTD-containing repeat homology <YW34>
F;3070-3114/Domain: LDL receptor WYTD-containing repeat homology <YW35>
F;3115-3157/Domain: LDL receptor WYTD-containing repeat homology <YW36>
F;3158-3201/Domain: LDL receptor WYTD-containing repeat homology <YW37>
F;3202-3242/Domain: LDL receptor WYTD-containing repeat homology <YW38>
F;3243-3285/Domain: LDL receptor WYTD-containing repeat homology <YW39>
F;3295-3331/Domain: EGF homology <EG13>
F;3335-3370/Domain: LDL receptor ligand-binding repeat homology <LDLL>
F;3375-3409/Domain: LDL receptor ligand-binding repeat homology <LDLM>
F;3414-3449/Domain: LDL receptor ligand-binding repeat homology <LDLN>
F;3454-3490/Domain: LDL receptor ligand-binding repeat homology <LDLO>
F;3495-3532/Domain: LDL receptor ligand-binding repeat homology <LDLP>
F;3537-3571/Domain: LDL receptor ligand-binding repeat homology <LDLQ>
F;3576-3610/Domain: LDL receptor ligand-binding repeat homology <LDLR>
F;3614-3648/Domain: LDL receptor ligand-binding repeat homology <LDLS>
F;3655-3691/Domain: LDL receptor ligand-binding repeat homology <LDLT>
F;3696-3732/Domain: LDL receptor ligand-binding repeat homology <LDLU>
F;3742-3777/Domain: LDL receptor ligand-binding repeat homology <LDLV>
F;3786-3823/Domain: EGF homology <EG14>
F;3829-3861/Domain: EGF homology <EG15>
F;3869-3912/Domain: LDL receptor WYTD-containing repeat homology <YW40>
F;3913-3970/Domain: LDL receptor WYTD-containing repeat homology <YW41>
F;3945-4545/Domain: alpha-2-macroglobulin receptor 85K chain #status predicted <85K>
F;3971-4013/Domain: LDL receptor WYTD-containing repeat homology <EXT>
F;4014-4057/Domain: LDL receptor WYTD-containing repeat homology <YW42>
F;4058-4100/Domain: LDL receptor WYTD-containing repeat homology <YW43>
F;4101-4143/Domain: LDL receptor WYTD-containing repeat homology <YW44>
F;4152-4183/Domain: EGF homology <EG16>
F;4201-4232/Domain: EGF homology <EG17>
F;4237-4268/Domain: EGF homology <EG18>
F;4273-4304/Domain: EGF homology <EG19>
F;4309-4340/Domain: EGF homology <EG20>
F;4345-4375/Domain: EGF homology <EG21>
F;4378-4409/Domain: EGF homology <EG22>
F;4422-4445/Domain: transmembrane #status predicted <TMM>
F;4446-4545/Domain: intracellular #status predicted <INT>
F;167,2999/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F;2959/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
F;4076,4126,4279/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.5%; Score 2587.7; DB 1; Length 4545;
Best Local Similarity 10.0%; Pred. No. 2e-52;
Matches 431; Conservative 59; Mismatches 13; Indels 3813; Gaps 386;
QY 1 MS-----I-----SSDE-----V-----N-F-L-----11-
Db 89 MSRLNGIQDMGSGEGAHCRANCRMGQHCVCPTPTSGTCVNSFFLEADGKT 148
QY 12 -----VY-----R-----Y-LQ-----E-----18
Db 149 CKDFECSVYTCQLCTNTDGSFTGCGVEGYLQPDNRSCAKNEFVDRPPVLLIANQ 208
QY 19 -----SG-----FS-----H---SAF-T-----F 28
Db 209 NILATYLSGAQVSTFTPTSTQTTAMDFTSVANETVCVWVHVGDSAAQTOLCARMPLKGF 268
QY 29 GI-ESH---IS-----Q-----S-N-----I-----NG-----41
Db 269 -VDE-HTINISLSLHHVEQMAIDWLITGNFFVDDIDRIFVCNCRNGDTCVTLLDLELYNP 326

QY 42 --AL--VP-----P-----A- 47
Db 327 KGIALDPAMGVFTDYGQIPKVERCDMGQNRKTLVDSKIVFPHGITTLDVLSRLVTWAD 386
QY 48 A--LI--SIIQK-L-Q--Y--V-E-----A-----EVS-I----- 66
Db 387 AYLDIEVVDYEGKGRITIIQ-GILIEHLVLTVFENLYATNSDNANTQOKTSVIRVNR 445
QY 67 -N-E--D--GTL----- 72
Db 446 FNSTEYQVTVRDKGALHIYHQRRQPRVRSHACENDQYKPGGCSIDICLLANSHKARTC 505
QY 73 --F--DG--R-----R-----PIE-----SL 81
Db 506 RCRSGFSLGSDGSKCKPEHELFLVYKGRPGIIRGMGMGAKVPDEHMIPIENLMNPRAL 565
QY 82 -----S-LI-----D-----D-----AV-M-----PD 90
Db 566 DPHAETGFIYPADTTSYLIQRKIDGTERETILKGIHNVGVAVDMGDNLYWTDGPK 625
QY 91 --V-V--O--O--OA-----Y----- 99
Db 626 KTIISVARLEKAAQTRKTLIEGKMTHPRAIVVDPPLNGHMYTWDWEDPKDSRGRLEAWM 685
QY 100 --RD--K-----L--A--Q-----Q-----H 107
Db 686 DSHRDIFVTSKTVLWPNGLSLDIPAGRLYWVDAFYDRIETILLNGTDRKIVYEGPELNH 745
QY 108 A-----AAA-----A----- 112
Db 746 AFGLCHHGNYLFWTEYRSGSVYRLERGVAGAPPTVTLRSERPPIEIRMYDAHEQOQVT 805
QY 113 -----AA-----A-A-----AT--N-----Q----- 120
Db 806 NKCRVNGGSSICLATPGSRQACAEQVLDTDGTCLANPSYVPPQCPQCFACANN 865
QY 121 --Q-----G-----S--A-----K-----N--G----- 127
Db 866 RCIOERWKCDGNDCLDNDSEAPALCHOHTCPSPDRFKCENNRCIPNRWLCGDGNDGCNSE 925
QY 128 -E-N-T--A-----N-----G--E-----E----- 135
Db 926 DESNATCSARTCPNPFQSCASGRCPISWTCDLDDCGDRSDESASCAYPCTPPLTQFTC 985
QY 136 -NG-----A--H--T-----I----- 141
Db 986 NNGRCININWRCDNDGNSDEAGCSHSCSTQFKNSGRCPPEHWTCDGNDGCDYS 1045
QY 142 --AN--N-----HTDME-----V-----DGDV-----E----- 156
Db 1046 DETHANCTNQAATRPFGGCHSD--EFQCPLDGLCIPLRWRCGDTDCMDSDEKSCGVTH 1103
QY 157 I--PSN--K-A-V-----V-----L--R----- 166
Db 1104 VCDP-NVFKGCKDSARCISNAWVCDGSDCEDNSDENECALACRPPHPCANNTSVCLP 1162
QY 167 -----G-----H-----E--S-----E----- 171
Db 1163 PDKLCKDKDCGDSDEGELCDQCGLNNGCSHNSVAPGEGIVCSPLGMELGSDNHTC 1222
QY 172 -V-----F-I-CA--W-NP--V--S-D--LLV--S----- 186
Db 1223 QIQSACAKHLKCSQKQDONKFSVKCSYEGWVLEPDEGTCSRSLDPFKLFIIFSNRHEIR 1282
QY 187 -----G--S-----D-----S--T-A-----R-----I-- 195
Db 1283 IDLHKGDYSVLVPEGLANTIALDPHLSQSALYWTDAVEDKTYRKLNDNGALTSFEVVIQY 1342
QY 196 -----W--N-L-SE-N-----ST--SG-----PTO--L-----V 210
Db 1343 GLATPEGLAVDWIAGNIYWVESNLQIEVAKLDGTLRLLAGDI BHP-RAIALDPRDGI 1401
QY 211 L-----R--H-----C-----I-- 215

Db 1402 LPWTDWDASLPRIEAAASMGAGRRTIHRETGSGCANGLTVDYLEKRIILWIDARSDAIS 1461
QY 216 -----R--E-----GG----- 219
Db 1462 ARYDGSCHMEVLRGHEFLSHPPFAVTLYGGEVYVWDMRTNTLAKANKWTHNVTVVQRTNT 1521
QY 220 Q--D--V--PS-----N-----N-----KD----- 227
Db 1522 QPFDLQVYTHPSROPMAPNPCANGRGPCSHCLINYNRTVSWACPHMLKHLKNDTTCYE 1581
QY 228 -----N-SEG--T--L--LA-----T-----GS----- 244
Db 1642 RAFINGT-GVETVVSADLPNAHGLAVDMVSRNLFWTSYDNTKKQINVARLDGSPKNAVQ 1700
QY 245 -----Y-----DG--F-----A----- 249
Db 1701 GLEQPHGLVHPLRGKLYWTDGDNISMANMDGSNHTLLFSGQKGPVCLAIIDFESKLYWI 1760
QY 250 -----R-----I-----W-----T-K-DG-- 256
Db 1761 SSGNHTINRCNLGDSLEVIDTMRSQLGATALAIMGDKLWADQVSEKMGTCNKADGSG 1820
QY 257 -----N--LA-----S--T-----L 262
Db 1821 SVVLRNSTTLVMHMKVYDESIOLEHGTNPFCSVNNGDCSOLCLPTSETTSCMCTAGYSL 1880
QY 263 --CQ-----HKG-----P-----I-E-A----- 271
Db 1881 RSCQACEGVGSPLLSVHEGIRGIPLDPNDKSDALVPVSGTSLAVGIDFHAENDTIYVW 1940
QY 272 --L-----K--W--N--K-----KGN-----F-I--L----- 282
Db 1941 DMGLSTISRAKRDQWREDVVTNGIGRVEGIAVDWIAIGNIYWTDOGSFVIEVARLNGSFR 2000
QY 283 --SAGVDK-----T-----T-----I-----I-- 292
Db 2001 YVVISQGLDKPRAITVHPEKGYLFWTEWGHYPRIERSRLDGTERRVVLNVNVSISWPNGISV 2060
QY 293 -----W-DA-----H----- 296
Db 2061 DYOGGKLYWCDAEMDKIERIDLETGENREVLLSSNNMDPSVSVFEDFIWSDRTHANGS 2120
QY 297 -----TG----- 298
Db 2121 IKRGCKDNATDSVPLRTGTIGVQLKDIKVFNRDRQKGTNVCVANGGQQCLLYRGGQRA 2180
QY 299 -----E--A-----K----- 301
Db 2181 CACAHGMLAEDGASCREYAGYLLYSERTILKSIHLSDERNLNAPVOPFEDPEHMKNVIAL 2240
QY 302 -----QO----- 303
Db 2241 AFDYRAGTSPGTNRIFFSDIHFNGIQIINDCGSRITIVENVGSVEGLAYHRGWDTLYW 2300
QY 304 -----F-----P--F-----H----- 307
Db 2301 TSYTTSTITRHTVDTQTRPGAFERETVITWSGDDHPRAFVLDECONLMFMTNWNELHPSIM 2360
QY 308 --S-A-----P-----A-LD----- 313
Db 2361 RAALSGANVLTIEKDIRTPNGLAIDHRAEKLYFSDATLDKIERCEYDGSHRVILKSEP 2420
QY 314 -----V-----DW-----Q----- 317
Db 2421 VHPFGLAVGEHIFWTDWVRRAVORANKYVGSMDKLLRVDIPOQPMGIIVANDTNSCEL 2480
QY 318 S-----NN-----T-----F--A-----SC-STD--M--CI-- 331

Db 2481 SPKINNGGQDLLTHQGHVNCRCRGRILQEDFTCRVAVNSSCRAQDEFECANGECIS 2540
QY 332 ---H---V--- 333
Db 2541 FSUTCDGVSHCKDSDEKPSYCNRRCKTFRQNNRCVSNMLWNGVDYCGDGSDEIP 2600
QY 334 C-K---LG---Q- 338
Db 2601 CNKTACGVGEPRCDGSCIGNSSRCNFQVDCEDASDEMNCSATDCSSYFRLGVKGVLP 2660
QY 339 ---D---RP---K- 343
Db 2661 CERTSLYAPSWYCDGANDCGDYSDERDCPVKRCPLNYFACPSGRCPMSWTCDKED 2720
QY 344 ---T---F---Q--- 346
Db 2721 DCENGEDETHCNKFCSEAFECQNHRCISQWLCDGSDDCGSDGDEAAHCEGKTCGPSSF 2780
QY 347 ---G-H---T---N---E---V--- 352
Db 2781 SCFGTHVCVPERWLCDDKCDCTGDADESVTAGCLYNSTCDDREFMCONRLCIPKHVCDH 2840
QY 353 ---N---A---I---KW---D---P-T- 360
Db 2841 DRDCADGSDSPCEVPTCGNEFRCAANGRCCLSSQWECDEGENDCHDHSDEAPKNPCHTS 2900
QY 361 ---L---G---N---L 363
Db 2901 PEHKCNASSQFLCSSGRCAEALLCNGQDCGSDPERGCHVNECLSRKLSGCSQDCEDL 2960
QY 364 ---L---A---S---CS---D--- 369
Db 2961 KIGFKRCRCPFLKDDGRTCADLDECSFTFPCSQLCINTHSGYKCLCVGEYAPRGDPH 3020
QY 370 ---D---M---T-LK---I-W--- 376
Db 3021 SCAVTDDEEPLIFANRYLRKLNLDGNSVTLKQGLNNAVALAFDYREQMIYWTGTTQ 3080
QY 377 -SM--K---Q---D---N---C--- 383
Db 3081 GSMIRMHNGSNVQLHRTGLSNPDGLAVDWGGLYKCDKGRDTEVSKLNGAVRTVL 3140
QY 384 ---V---H---H--- 385
Db 3141 VSSGLEPRALVVDVQNGYLYWTWDGHSILGRIGMDGSGRSIIVDTKITWPNGLTVDYV 3200
QY 386 ---D---L-Q---AH- 390
Db 3201 TERIYADAREDYEFASLDGNSRHVVLSQDIPHFALTFLFYVYVWTDWETKSINRAHK 3260
QY 391 ---NK---EI--- 394
Db 3261 TTGANKTLTSLHRPMDLHVHFAHQDPVNPHPKYNNGGCSNLCILSPGGHKACAPT 3320
QY 395 -Y---T---IK-W---S---P--- 401
Db 3321 NFYLGGDRTCVSNCTASQFVKNDKCIPPWKKCDTDEDDCGDHSDEPPDCPEFKCRGQF 3380
QY 402 ---TG---P---G---TN---N- 408
Db 3381 QCSTGICTNPAFICDGDNDQDNDSEANCIDHVLCPSPQFKTNTNRCIPGIFRCNGQDNC 3440
QY 409 ---PN---ANL-M---L 415
Db 3441 GDGEDRDCPEVTCAPNQFQCSITKRCIPRVWVCDRDNHCVDGSDPEANCTQTCGVDEF 3500
QY 416 ---AS---A---S---F--- 420
Db 3501 RKDSDGRCIPARWKCDGDCDGSDFPKBEDERTCEPQFRCCKNNRCVGRWQCYDN 3560
QY 421 ---D-S-T---V---R--- 425
Db 3561 DCGDSDSEESCTPRPCSESEBFFCANGRCIAGRWKCDGHDCAADGSDGDEKCTPRCDMDQFQ 3620

QY 426 ---L-W---DVD---R--- 431
Db 3621 CKSGHCIPLRNFCADADCMGDSDEACGTGVRTCPDLDFQCNNTLCKPLAWKCDGEDDC 3680
QY 432 G---IC---I--- 435
Db 3681 GDSNDENPECARFICPPNRPFRCKNDRVCLMTGRQCQDVNCGDGTDEBCEPPTAQN 3740
QY 436 H---T---L--- 438
Db 3741 HCKDKKEFLCRNQCLSSRLRCNMFDDCGDSDDEDCSDPKLTSATNASMGDEARCV 3800
QY 439 -T-K---H---Q-E---P--- 444
Db 3801 RTEKAYACACRSGFHTVPGQCGQDINECLRFCTCSQLWNKPKGGHLCSCARNFMKTHNT 3860
QY 445 ---V-YSA---FSP---D---GR--- 455
Db 3861 CRAEGSEYQVLY-IADNBEIRSLF-PGHPHSAEQTFQGDSEVRIDAMDVHVKAGRVYT 3918
QY 456 ---Y-L---A---S---G---S---F--- 462
Db 3919 NMHTGTISYRSLPPAAPPTTSNRHRQIDRGVTHLNLISGLKMPRGIAIDWVAGNVYWTDS 3978
QY 463 -D---K-C---V-H-I---W-N--- 470
Db 3979 GRDVEIAQMKGNRKTILSGMIDEPHAIIVDPLRGTMYSWDMGNHPKIEAAMDGLRE 4038
QY 471 T---Q-TGAL-V-HS---Y---R-GT---G--- 484
Db 4039 TLVQDNIQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPIVAADSKRGLSHPFSD 4097
QY 485 ---G---I---F---EV---C 490
Db 4098 VFEDYIYGVTYINNRVFKHKFGHSPLYNLTLGSLSHASDVVLVHQHKQPEVTNPCRKKC 4157
QY 491 -W---N---N---A---A- 494
Db 4158 EWLCLLSPSGVCTCPNGKRLDNGTCVPVPSPTPPDARPPGCTTLQCFNGSGCFLNARR 4217
QY 495 ---GDK---V---G---AS---A--- 502
Db 4218 QPKRCQPRYTGDKCELDQWECYCHNGGTCAASPSGMPTRCPTGFTGPKCTAQVCAGYC 4277
QY 503 ---S---DGS---V--- 507
Db 4278 SNNSTCTVNGNQPCRCPLPGFLGDRQYRQSGPCFENFGTCQMAADGSRQCRCTVYFEG 4337
QY 508 -C-V---L---DL---R---K 514
Db 4338 PRCEVKNCSRLOGACVNVNQTGDTCTNCTDGRVAPSLCTCIDHCSNGSGSCTWNSK 4393

RESULT 9

S02392

N;Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C;Species: Homo sapiens (man)

C;Date: 14-Aug-1998 #sequence revision 14-Aug-1998 #text_change 09-Jul-2004

C;Accession: S02392; S30027; I37998; A39210; S12538

R;Herz, J.; Hamann, U.; Rogne, S.; Myklebost, O.; Gausepohl, H.; Stanley, K.K.

EMBO J. 7, 4119-4127, 1988

A;Title: Surface location and high affinity for calcium of a 500-kd liver membrane protein

A;Reference number: S02392; MUID:89210795; PMID:3266596

A;Accession: S02392

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-4544 <HER>

A;Cross-references: UNIPROT:Q07954; EMBL:X13916; MID:g34338; PIDN:CAA32112.1; PID:g34339

R;Kristensen, T.

submitted to the EMBL Data Library, October 1990

A;Reference number: S30027

A;Accession: S30027
 A;Molecule type: mRNA
 A;Residues: 3275-3864 <KRI>
 A;Cross-references: EMBL:X55077
 R;Harz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.
 EMBO J. 9, 1769-1776, 1990
 A;Title: Proteolytic processing of the 600 kd low density lipoprotein receptor-related protein
 A;Reference number: S12538; MUID:90269210; PMID:2112085
 A;Contents: annotation; site of proteolytic cleavage
 R;Kutt, H.; Herz, J.; Stanley, K.K.
 Biochim. Biophys. Acta 1009, 229-236, 1989
 A;Title: Structure of the low-density lipoprotein receptor-related protein (LRP) promotes
 A;Reference number: I37998; MUID:90089395; PMID:2597675
 A;Accession: I37998
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-11 <RES>
 A;Cross-references: EMBL:X15424; NID:934408; PIDN:CAA33464.1; PID:934409
 R;Strickland, D.K.; Ashcom, J.D.; Williams, S.; Burgess, W.H.; Migliorini, M.; Argaves,
 J. Biol. Chem. 265, 17401-17404, 1990
 A;Title: Sequence identity between the alpha2-macroglobulin receptor and low density lip
 A;Reference number: A39210; MUID:91009181; PMID:1698775
 A;Accession: A39210
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 150-166;234-238,'X',240-245,'X',247-252;'G',686-695;902-916;1096-1109;'S',17
 C;Genetics:
 A;Gene: GDB:LRP1; APR: LRP; A2WR
 A;Cross-references: GDB:111694; OMIM:107770
 A;Map position: 12q13.1-12q13.3
 C;Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated
 ciated protein (see PIR:A39875).
 C;Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
 C;Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glycopro
 F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-3943/Product: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>
 F;27-64/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F;72-108/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;115-148/Domain: EGF homology <EG1>
 F;154-188/Domain: EGF homology <EG2>
 F;198-239/Domain: LDL receptor WYTD-containing repeat homology <YW01>
 F;240-281/Domain: LDL receptor WYTD-containing repeat homology <YW02>
 F;292-334/Domain: LDL receptor WYTD-containing repeat homology <YW03>
 F;335-378/Domain: LDL receptor WYTD-containing repeat homology <YW04>
 F;379-420/Domain: LDL receptor WYTD-containing repeat homology <YW05>
 F;421-468/Domain: LDL receptor WYTD-containing repeat homology <YW06>
 F;478-519/Domain: EGF homology <EG3>
 F;571-613/Domain: LDL receptor WYTD-containing repeat homology <YW07>
 F;614-659/Domain: LDL receptor WYTD-containing repeat homology <YW08>
 F;660-710/Domain: LDL receptor WYTD-containing repeat homology <YW09>
 F;711-752/Domain: LDL receptor WYTD-containing repeat homology <YW10>
 F;753-799/Domain: LDL receptor WYTD-containing repeat homology <YW11>
 F;807-842/Domain: EGF homology <EG4>
 F;854-890/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F;895-931/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F;936-971/Domain: LDL receptor ligand-binding repeat homology <LDL5>
 F;976-1011/Domain: LDL receptor ligand-binding repeat homology <LDL6>
 F;1015-1051/Domain: LDL receptor ligand-binding repeat homology <LDL7>
 F;1062-1097/Domain: LDL receptor ligand-binding repeat homology <LDL8>
 F;1104-1140/Domain: LDL receptor ligand-binding repeat homology <LDL9>
 F;1145-1182/Domain: LDL receptor ligand-binding repeat homology <LDL10>
 F;1185-1221/Domain: EGF homology <EG5>
 F;1227-1261/Domain: EGF homology <EG6>
 F;1269-1308/Domain: LDL receptor WYTD-containing repeat homology <YW12>
 F;1309-1355/Domain: LDL receptor WYTD-containing repeat homology <YW13>
 F;1356-1398/Domain: LDL receptor WYTD-containing repeat homology <YW14>
 F;1399-1445/Domain: LDL receptor WYTD-containing repeat homology <YW15>
 F;1446-1488/Domain: LDL receptor WYTD-containing repeat homology <YW16>
 F;1489-1531/Domain: LDL receptor WYTD-containing repeat homology <YW17>
 F;1540-1578/Domain: EGF homology <EG7>
 F;1583-1626/Domain: LDL receptor WYTD-containing repeat homology <YW18>
 F;1627-1669/Domain: LDL receptor WYTD-containing repeat homology <YW19>
 F;1670-1713/Domain: LDL receptor WYTD-containing repeat homology <YW20>

F;1714-1753/Domain: LDL receptor WYTD-containing repeat homology <YW21>
 F;1754-1796/Domain: LDL receptor WYTD-containing repeat homology <YW22>
 F;1797-1846/Domain: LDL receptor WYTD-containing repeat homology <YW23>
 F;1850-1886/Domain: EGF homology <EG8>
 F;1934-1976/Domain: LDL receptor WYTD-containing repeat homology <YW24>
 F;1977-2019/Domain: LDL receptor WYTD-containing repeat homology <YW25>
 F;2020-2063/Domain: LDL receptor WYTD-containing repeat homology <YW26>
 F;2064-2105/Domain: LDL receptor WYTD-containing repeat homology <YW27>
 F;2106-2151/Domain: LDL receptor WYTD-containing repeat homology <YW28>
 F;2159-2194/Domain: EGF homology <EG9>
 F;2199-2241/Domain: LDL receptor WYTD-containing repeat homology <YW29>
 F;2253-2294/Domain: LDL receptor WYTD-containing repeat homology <YW30>
 F;2344-2388/Domain: LDL receptor WYTD-containing repeat homology <YW31>
 F;2389-2429/Domain: LDL receptor WYTD-containing repeat homology <YW32>
 F;2430-2473/Domain: LDL receptor WYTD-containing repeat homology <YW33>
 F;2482-2517/Domain: EGF homology <EG10>
 F;2524-2561/Domain: LDL receptor ligand-binding repeat homology <LDL1B>
 F;2566-2600/Domain: LDL receptor ligand-binding repeat homology <LDL1C>
 F;2605-2639/Domain: LDL receptor ligand-binding repeat homology <LDL1D>
 F;2652-2688/Domain: LDL receptor ligand-binding repeat homology <LDL1E>
 F;2696-2730/Domain: LDL receptor ligand-binding repeat homology <LDL1F>
 F;2734-2769/Domain: LDL receptor ligand-binding repeat homology <LDL1G>
 F;2774-2812/Domain: LDL receptor ligand-binding repeat homology <LDL1H>
 F;2818-2853/Domain: LDL receptor ligand-binding repeat homology <LDL1I>
 F;2858-2897/Domain: LDL receptor ligand-binding repeat homology <LDL1J>
 F;2904-2939/Domain: LDL receptor ligand-binding repeat homology <LDL1K>
 F;2944-2980/Domain: EGF homology <EG11>
 F;2986-3021/Domain: EGF homology <EG12>
 F;3029-3068/Domain: LDL receptor WYTD-containing repeat homology <YW34>
 F;3069-3113/Domain: LDL receptor WYTD-containing repeat homology <YW35>
 F;3114-3156/Domain: LDL receptor WYTD-containing repeat homology <YW36>
 F;3157-3200/Domain: LDL receptor WYTD-containing repeat homology <YW37>
 F;3201-3241/Domain: LDL receptor WYTD-containing repeat homology <YW38>
 F;3242-3284/Domain: LDL receptor WYTD-containing repeat homology <YW39>
 F;3294-3330/Domain: EGF homology <EG13>
 F;3334-3369/Domain: LDL receptor ligand-binding repeat homology <LDL1L>
 F;3374-3408/Domain: LDL receptor ligand-binding repeat homology <LDL1M>
 F;3413-3448/Domain: LDL receptor ligand-binding repeat homology <LDL1N>
 F;3453-3489/Domain: LDL receptor ligand-binding repeat homology <LDL1O>
 F;3494-3531/Domain: LDL receptor ligand-binding repeat homology <LDL1P>
 F;3536-3570/Domain: LDL receptor ligand-binding repeat homology <LDL1Q>
 F;3575-3609/Domain: LDL receptor ligand-binding repeat homology <LDL1R>
 F;3613-3647/Domain: LDL receptor ligand-binding repeat homology <LDL1S>
 F;3654-3690/Domain: LDL receptor ligand-binding repeat homology <LDL1T>
 F;3695-3731/Domain: LDL receptor ligand-binding repeat homology <LDL1U>
 F;3741-3776/Domain: LDL receptor ligand-binding repeat homology <LDL1V>
 F;3785-3822/Domain: EGF homology <EG14>
 F;3828-3860/Domain: EGF homology <EG15>
 F;3868-3911/Domain: LDL receptor WYTD-containing repeat homology <YW40>
 F;3912-3969/Domain: LDL receptor WYTD-containing repeat homology <YW41>
 F;3944-4544/Product: alpha-2-macroglobulin receptor 85K chain #status predicted <EXT>
 F;3944-4544/Domain: 85K chain extracellular #status predicted <EXT>
 F;3970-4012/Domain: LDL receptor WYTD-containing repeat homology <YW42>
 F;4013-4056/Domain: LDL receptor WYTD-containing repeat homology <YW43>
 F;4057-4099/Domain: LDL receptor WYTD-containing repeat homology <YW44>
 F;4100-4142/Domain: LDL receptor WYTD-containing repeat homology <YW45>
 F;4151-4182/Domain: EGF homology <EG16>
 F;4200-4231/Domain: EGF homology <EG17>
 F;4236-4267/Domain: EGF homology <EG18>
 F;4272-4303/Domain: EGF homology <EG19>
 F;4308-4339/Domain: EGF homology <EG20>
 F;4344-4374/Domain: EGF homology <EG21>
 F;4377-4408/Domain: EGF homology <EG22>
 F;4421-4444/Domain: transmembrane #status predicted <TM>
 F;4445-4544/Domain: intracellular #status predicted <INT>
 F;166,2998/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
 F;2958/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
 F;4075,4125,4278/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.5%; Score 2586.1; DB 1; Length 4544;
 Best Local Similarity 10.0%; Pred. No. 2,2e-52;
 Matches 432; Conservative 65; Mismatches 8; Indels 3809; Gaps 387;

QY 1 MS-----I-----SSDE-----V-----N-F-L----- 11
Db 88 MSRLCNGVQDCMGDSDEGPHCRELOGNCRSLGQCHHCVPITLDGPTCYCNSSFOLQADGKT 147
QY 12 -----VY-----R-----Y-LQ-----E----- 18
Db 148 CKDFDECSVYGTCSQLCTNTDGSFICGCVGYLLQPDNRSCKAKNEPVDPPVLLIANSQ 207
QY 19 -----SG-----FS-----H-SAP-T-----F 28
Db 208 NILATYLSGAQVSTIPTSTROTMTAMDFSYANETVCVHVHVGDSAAQTOLKCARMPGLKGF 267
QY 29 GI-ESH--IS-----Q-----S-N-----I-----NG----- 41
Db 268 -VDE-HTINISLSLHVQWQAIWLTGNFYFVDIDRIFVCVRNGDTCVTLLDLELYNP 325
QY 42 --AL-----VP-----P-----A- 47
Db 326 KGIALDPAMGKVFTDYGOIPKVERCDMDGQNETKLVDSKIIVPHGITLTLVSLVYVWAD 385
QY 48 A---LI-----SIIQK-L-Q-Y---V-E-----A-EVS-I----- 66
Db 386 AYLDYIEVDYEGKGTIIQ-GILIEHLYGLTVFENLYATNSDNANAQOKTSVIRVR 444
QY 67 -N-E-----D-GTL----- 72
Db 445 FNSTEQVTVTRDKGALHIYHORRPRVRSHACENDQYKPGGCGSDICLLANSHKARTC 504
QY 73 --F-----DG-----R-----PIE-----SL 81
Db 505 RCRSGFSLGSGKCKPEHELFVYKGRPGIIRGMGMGAKVPDEHMIPIENLMNPRAL 564
QY 82 -----S-LI-----D-----AV-M-----PD 90
Db 565 DFHAETGIYFADTTSYLIQKIDGTERITLKGDIHNVGAVADMGMNLYWTDGPK 624
QY 91 -V-V-----QTR-----Q-----QA-----Y----- 99
Db 625 KTSVARLEKAAQRTKTLIBGKWTHPRAIIVDPLNGHMYTWDWEDPKDSRRGLERAWM 684
QY 100 --RD-----K-----L-----A-----Q-----Q-----H 107
Db 685 DGSHRDI FVTVKTLVWPNGLSLDIPAGRLYVWDAFYDRIETILLNGTDRKIVYEGPELNH 744
QY 108 A-----A-----A-----AA----- 111
Db 745 AFGLCHHGNYLFWTEYRSGSVYRLERGVGGAPPTVTLRLSRPPIFEIRMYDAQOQVGT 804
QY 112 -----AA-----A-A-----A-AT-----N-----Q----- 120
Db 805 NKCRVNGGSSSLCLATPGSRQCAEDQVLDADGVTCLANPSVPPQCPQGEFACANS 864
QY 121 --Q-----G-----S-----A-----K-----N-G----- 127
Db 865 RCIQERNKCDGNDCLNDSDEAPALCHOCTPCSPDRFCENNRCPINRWLDCGNDGCNSE 924
QY 128 -E-N-T-A-----N-----G-----E-----B----- 135
Db 925 DESNATCSARTCPFNQFSCASGRCPISWTCDLDDCGDRSDSASCAYTCFPLTQFTC 984
QY 136 -NG-----A-----H-----T-----I----- 141
Db 985 NNGRCININWRCDNDCGNSDEAGCSHSSSTQPKNSGRCPICEHWTCDGNDGCDYS 1044
QY 142 -----AN-N-----HTD-----M-----M-E-----V- 151
Db 1045 DETHANCTNQATRPFGCHTDFQCRDLGLCIPLRWRCGDGTDGMDSSDEKSCGVTHVC 1104
QY 152 D-----G-D-----V-----E-I-----PS-----NKAV----- 163
Db 1105 DPSVFGKDSARCISKAWCDGNDGDCEDNDSDEENCSLACRPPSHPCANNITSVCLPPDK 1164
QY 164 V-----L-R-----G-H-----E-----S-----E-----V- 172

Db 1165 LCDGNDCCDGSDEGELCDQCCLNNGCGSHNCVAPGEGIVCSPGLMBELGPDNHTCQIQ 1224
QY 173 -----F-I-CA-----W-----N----- 178
Db 1225 SYCAKHLKCSQKCDQNFVKSCCYEGWVLEPDGSCRSRLDPFKFPIIFSNRHRIRIDL 1284
QY 179 -----P-----V-----S-----D-----LL----- 184
Db 1285 HKGDYSLVPLGRNIALDFHLSQALYWTVDVVEDKIYRGKLLDNGALTSFEVVIQYGLA 1344
QY 185 -----V-S-----GS-----GD-----S-----T 192
Db 1345 TPEGLAVDWIAGNIYWVESNLQDIEVAKLDGLTTLTLAGDIEHPRAIALPRDGLIFWT 1404
QY 193 --A--RI-----W-N----- 197
Db 1405 DWDASLPRIEAAASMSGAGRRTVHRETGGGWPNGLUTVDYLEKRIILWIDARSDAIYSARYD 1464
QY 198 -----LS-----E-----NST-----SG----- 205
Db 1465 GSGHMEVLRGHEFLSHPPFAVTLYGGEVYWDHRTN-TLAKANKWTGHNVTVVQRTNQPF 1523
QY 206 --PT-----Q-----LV-L-R-----H-----C----- 214
Db 1524 DLQVYHPSRQPMAPNCPCEANGGQPCSHLCLINRYNRTVSCACPHLMKHKDNTTCYEFKK 1583
QY 215 -----I-----RE----- 217
Db 1584 FLYARQMEIRGVLDAPYNYIISFTVPDINVTVDYDAREQRYVSDVRTQAIKRAF 1643
QY 218 --G-G-----QDVPSNK-----D-V-----TS-----LD----- 232
Db 1644 INGTGVTVSADLP-NAHGLAVDWYSRNLFWTSYDTNKKQINVARLDGSKFNAVVQGLE 1702
QY 233 -----W-----N-S-----EG-----TLL-A-TG-----S-Y----- 245
Db 1703 QPHGLVVHPLRGKLYWTDGDNISMANMDGSRNTLLFSQKGPVGLAIDFPESKLYWISSG 1762
QY 246 -----D-GF-----A-R-----I-----W-----T-K-DG----- 256
Db 1763 NHTINCNLDGSGLEVIDAMRSOLGKATALAIMGDKLWADQVSEKMGTCCKADGSGSVV 1822
QY 257 --N---LA-----LA-----EG-----TLL-A-TG-----S-T-----L-G 263
Db 1823 LRNSTTLVMMKMYDESIIQLDHKGTPNCSVNNNGDCSQLCLPTSETTRSCMCTAGYSLRSG 1882
QY 264 Q-----HKG-----P-----I-F-A----- 271
Db 1883 QQACEGVGSFLLYSVHEGIRGIPLDPNKSDALVPVSGTSLAYGIDFHAENDTIYVWDMG 1942
QY 272 L-----K-----W-----N-----KGN-----P-I-----L----- 282
Db 1943 LSTISAKRQDQWREDVVTNGIRGVEGIAVDWIAAGNIYWTDOQDFVIEVARLNGSFYVV 2002
QY 283 -SAGVDK-----T-----T-----I-----I----- 292
Db 2003 ISQGLKXPRAITVHPKEGYLFWTEWGOYPRIERSLDGTERVVLVNVVSIWPNGISVDYQ 2062
QY 293 --W-DA-----H----- 296
Db 2063 DGKLYWCARDARTKIERIDLETGENREVVLSSNNMDMFSVSVFEDFIYWSRTHANGSIKR 2122
QY 297 -----TG----- 298
Db 2123 GSKDNATDSVPLRTGLGVQLKDIKVFNRDRQKNTNVCANGCGQQLCLYGRGQRACAC 2182
QY 299 -----E-A-----K-----Q-----Q-----F- 304
Db 2183 AHGMLAEDGASCREYAGYLYSERTILKSTHLSDERNLNAPVQPFEDPEHMKNVIALAFD 2242
QY 305 --P-----F-H-----S-----A----- 309

Db 2243 YRAGTSGTNRIPFFSDIHFGNIQQINDGSRRTITVENVSGVEGLAYHRGMDLYWTSY 2302
QY 310 -----P-AL-----D-----V-D-----W-Q----- 317
Db 2303 TTSTITHTVDTQTRPFAFERETVITMSGDDHPRAFLVDEQNLMFTWNEQHPSIMRAA 2362
QY 318 -S-N-----N-----T----- 321
Db 2363 LSGANVLTIKEDIRTPNGLAIDHRAEKLKFSATLKDIERCEYDGHRYVLKSEBPVHP 2422
QY 322 -----F-----A-----SC-S-- 326
Db 2423 FGLAVYGEHIFWTDWVRRAVQRANKHVGSNMKLLRVDIPOQPMGIIIAVANDTNSCSP 2482
QY 327 -----T-----DM-----CI----- 331
Db 2483 RINNGGQDLCLLTHQHVNCSCRGRILODDLTCTRAVNSSCRAQDEFECANGECINFSL 2542
QY 332 -----H-----V-----C-K 335
Db 2543 TCDGVPHCKDSDEKPSYCNRRCKKTFRQCSNGRCVSNMLWCNMGADDCGDSDEIPCNK 2602
QY 336 -----F-----Q----- 338
Db 2603 TACGVGEFRCDGTGICGNSRCNQFVDCEDASDEMNCSATDCSSYFRLGVKGLFQPCER 2662
QY 339 -----D-----RP-----I-----K----- 343
Db 2663 TSLCYAPSVWCDGANDGSDYDRDCPGVVRPCPLNYFACPSGRCI PMSWTCKEDDCE 2722
QY 344 -----T-----F-----Q-----G-----H----- 348
Db 2723 HGEDETHCNKFCSEAFQECQNHRCISKQWLCDGSDGDEAAHCEGKTCGPSFSFSCP 2782
QY 349 -T-----N-----EV-----N-A-I-K----- 356
Db 2783 GTHVCVPERWLCDDKADGADESAAGCLYNSTCDDREFMQNRQCI PKHFVCDHRRD 2842
QY 357 -----W-----D-----P-T----- 360
Db 2843 CADGSDSECEYPTCGPSEFRANCGRCLSSRQWECDEGNDCHQDSEAPKPHCTSP 2902
QY 361 -----G-----N-----L----- 363
Db 2903 KCVASSQFLCSSGRCAEALLCNGQDDCGDSSDERGCHINECLSRKLSGCSQDCEDLKIG 2962
QY 364 -----L-----A-----S-----CS-----D----- 369
Db 2963 FKCRCPGFLKDDGRTCADVDCESTTFPCSQRCINTHGSYKCLCEGYAPRGDPHSCK 3022
QY 370 -----D-----M-----T-LK-----I-W-----SM 378
Db 3023 AVTDEPFLIFANRYLRLKNDGNSYLLKQGLNNAVALDFYREQMIYTDVTTQGS 3082
QY 379 -----K-----Q-----D-----N-----C----- 383
Db 3083 IRRMHLNGSVQVLRHTLSNPDLGAVDWGGLNYCCKGRDTIEVSKLNGAVRTVLVS 3142
QY 384 -----V-----H----- 385
Db 3143 GLREPRALVVDVQNGLYLWTDGHDHSLIGRMGSSRSVIVDTKITWPNGLTLDYVTER 3202
QY 386 -----D-----L-Q-----AH----- 390
Db 3203 IYADAREDIIEFASLDGSRHVLVSDIPHALTLFEDYVYVTDWETKSNRAHKTG 3262
QY 391 -NK-----EI-----Y 395
Db 3263 TNKTLI STLHRPMDLHVHALRQDPVPHPCVKNNGGCSNCLLSPGGGHKACACPTNFY 3322
QY 396 -----T-----IK-W-----S-----P----- 401
Db 3323 LGSGRTCVSNCTASQFVCKNDKCI PFVWKNKCDTDDCGHSDPEPPDCPBFKCRPGQFQCS 3382

RESULT 10
A88852
protein unc-22 [imported] - Caenorhabditis elegans

QY 402 TG-----P-----G-----TN-----N----- 408
Db 3383 TGICTNPAFICDGDNDQDNDSEANCDIHVCLPSQFKCTNTNRCIPGIPRCNGODNCGDG 3442
QY 409 -----PN-----ANL--M----- 414
Db 3443 EDERDCPEVTCAPNQFQCSITTKRCIPRVVWCDRDNDCVDSGDEPANCQTGCVDEPRCK 3502
QY 415 -----L-A----- 416
Db 3503 DSGRCIPARWKDGEDDCGDSDEPKEDCETCEPYQFRCKNNRCVPGRWQCDYDNDG 3562
QY 417 -----S-----A-----F----- 420
Db 3563 DNSDEBESCTRPCSESEFSCANGRCIAGRWKCDGDDHDCADGSDKDEKCTPCRDMDQFOCKS 3622
QY 421 -----D-S-----T-VR-----L-W----- 427
Db 3623 GHCIPLRWRCADADCMGDSDEEACGTGVRTCPDEFQCNNTLCKPLAMKCDGEDDCGDN 3682
QY 428 -D-----V-----DRGIC-I-----HTLTK 440
Db 3683 SDENPECARFVCPNRPFRCKNDR-VCLMIGRQCDGTNCGDGTDEDECEPPTAHT-T- 3739
QY 441 H-----Q-----E-----P-----V 445
Db 3740 HCKDKKEFLCRNQRLSSSLRCNMFDDCGDSEEDCSIDPKLTSCATNASICGDEARCV 3799
QY 446 -----Y-----S-----V-----A-F----- 450
Db 3800 RTEKAAYCACRGFHTVPGQCGDINECLRFQTCSQLCNNTKGGHLCSCARFMKTHNT 3859
QY 451 -----S-----P-----D-----CR----- 455
Db 3860 CKAGSEYQVLYIADNNEIRSLFPFPHSAYEQAFQGDSESVRIDAMDVHVKAGRVYTNW 3919
QY 456 -----Y-L-----A-----S-----G-----S-----F----- 462
Db 3920 HTGTISYRSLPAAAPPTTSNRHRRQIDRGVTHLNLISGLKMPRGIAIDVAGNVWYDTSGR 3979
QY 463 D-----K-C-----V-H-I-----W-N-----T- 471
Db 3980 DVIEVAQMGKGNKRTLISGMIDEPHAI VVDPLRGTWYSDWGNHPIETAAMDGTLRETL 4039
QY 472 -----Q-----TGAL-V-HS--Y-----R-GT-----G----- 484
Db 4040 VQDNTQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPIVAADSKRGLSHPSIDVF 4098
QY 485 -----G-----I-F-----EV-----C-W 491
Db 4099 EDYIYGVTVNNRVFKIHKFHSPLVNLTGGLSHASDVVLYHQHKQPEVTNPNCDRKKCEW 4158
QY 492 -----N-----A-----A----- 494
Db 4159 LCLLSPSGPVCTCPNGKRLDNGTCVPVPSPPTPPDAPRGTCNLQCFNGSGCFLNARQ 4218
QY 495 -----GDK-----V-G--A- 500
Db 4219 KCRCPRYTGDRCCLDQCEWHEHCRNGGTCAASPSGMPCTCPTGFTGPKCTQQVCAGYCAN 4278
QY 501 -----S-----ASDGS-----K 514
Db 4279 NSTCTVNGNQPCRCPLPGFLGDRCOYRQCSGYCENFGTCQMAADSGRQRCRTAYFEGSR 4338
QY 507 --V-----CVL-----DL-----R-----K 514
Db 4339 CEVKNKSRCLGACVVKQSGDVTCTDGRVAPSLCTCVGHCSNGSGSCTMNSK 4392

178 QY -----N-----PVS-----D-----L----- 183
2643 Db LKPHIDRALKNLTIRKAGQISFDVPVSGEPAPVTWHDNREINGRVLNPEYQS 2702
184 QY -----SG-----S-----GD----- 190
2703 Db KLVVQMERGDSGFTTIKAVNAGEDATVKNIDKPTSPNGPLDVSVDVHGHTLNR 2762
191 QY -----S-----TA-----R----- 194
2763 Db APDDGGIPIENTVIEKYDTASGRWPAKAVAGDKTTAVVDGLIPGHEYKFRVAANAEG 2822
195 QY -----I----- 195
2823 Db ESDPLEFTGTTAKDPDKPKGTNAPEITDWDKHDVLEWKPPANDGGAPIEYVVEMKD 2882
196 QY -----W-----NL-----S-----E----- 200
2883 Db EFSFFWNDVAHPAGOTNATVGNLKEGSKYEFIRAKNKAGLGDPSDSASAVAKARNVP 2942
201 QY -----NS-----TSG-----PT-----Q-----L-----V 210
2943 Db VIDRNSIQEIKVKAGQFSLNIPVSGEPTTIITWTEGTPVESDRMKLNEDGKTFHV 3002
211 QY -----LR-----H-----C----- 214
3003 Db KRALRSDTGTIYIIKAENENGDTAEVKVTVLDHPSPRGLDVTNIVKDCDLAWKEPED 3062
215 QY -----I----- 215
3063 Db DGAIEISHYVIEKQDAATGRWTAGESKOTNFHVDLITQGEYKFRVKA VNRHGDSDPLE 3122
216 QY -----RE----- 217
3123 Db AREALIAKOPFDRADKPGTPEIVDKHADLKWTPPADGGAPIEGYLVEMTSPSGDW 3182
218 QY -----G----- 218
3183 Db PAVTVAGBELTATVDGLKPGQTYQFRVKALNKAGESTPSPSRTMVAKPRHLPKINRDM 3242
219 QY -----GO-----DV-----P----- 223
3243 Db FVAQVRKAGQTLNFDVNVEGEPAPKIEWFLNGSLSSGGNTHIDNNTKLTTKSTAR 3302
224 QY -----S-----N-----KDV-----S-----LD----- 232
3303 Db ADGKYIVATNESGKDEHVDVNIILDI PGAPGLRHKDIITKESVVLKWDPLDDGGSP 3362
233 QY -----W-----N-----SS-----GT-----L----- 239
3363 Db ITNYVVEKQDGRWVPCGETSPTSLSKVNKLSGHEYKFRVKA VNRQGTSAPLTSDHAIV 3422
240 QY -----L-----A-----T-----G-----SY-----D 246
3423 Db AKNPFDEPDAPTDTVPVMDKHDVLEWKPPANDGGAPIDAYIVEKKDKFGDWECARVD 3482
247 QY -----G-----F----- 251
3483 Db GKTTKATADNLTPGETYQFRVKA VNRKAGPKPSDPTGNVVAKPRRMAPKLNLAGLDR 3542
252 QY -----W-----TKD-----G-----NL-----A-----S-----T-----L----- 262
3543 Db KAGTPIKLDIAPEGEPAAPVAKWKANDATIDTGARADVTTPTSSAIIHIFSAVRGDTGVYK 3602
263 QY -----G-----Q-----HK-----G-----P-----I 269
3603 Db IIVENEHGKDTACQCNVTLVDVPGTPEGPLKIDBIHKEGCTLNWKPPPTDNGGTDVLHYVE 3662
270 QY -----T-----F-----A-----L----- 272
3663 Db KMDTSRGTWQEVGTFPDCTAKVNKLVPGEYAFRVRKAVNQLQGESKPLEAEPIIAKNQFD 3722

273 QY -----KWN-----KK-----G----- 278
3723 Db VPDVDPKPEVTDWDKDRIDIKNPTANNNGGAPVTGYIVEKKEGSAIWEAGTPTGTPFS 3782
279 QY -----N-----P-----IL-----S-----AG----- 285
3783 Db ADNLKPGVEYFRVIAVNAAGSPSDPTDQITKARYLKPKILTASRKIKIRAGFTHNL 3842
286 QY -----VD-----K-----TT----- 290
3843 Db EVDFIGAPDPTATWTVGDSGAALAPPELLVDKSGSTTSIFPPPSAKRADSGNYKLVKKNELG 3902
291 QY -----I-----I-----W-----D----- 294
3903 Db EDEAI FEVIVQDRPSAPEGLEVDVTKDSCVLNWKPPKDDGAEISNYVVEKRDTKTNT 3962
295 QY -----A-----H-----T-----G-----E-----AKOOF----- 304
3963 Db WVPVSFAVVTGTSITVPKLTGEGHEYFRVMAENTFGRSDSLNTDEPVLAKDPFGTGGPKGR 4022
305 QY -----P-----F----- 306
4023 Db PEIVOTDNHDIDIKWDPDRDNGSGSPVDHYDIERKDAKTGRWIKWNTSPVOGTAFSDTRVQ 4082
307 QY -----H-----SA-----P-----A-----LD-----V----- 314
4083 Db KGHTEYRVRVAVNKAQPGQPSDSSAAATAKPMHEAPKFDLDLDGKEFRVKAGEPLVITIP 4142
315 QY -----D-----W-----Q-----S----- 318
4143 Db FTASPOQDISWTKEGGKPLAGVETDTSQTKLVIPSTRSSDGPVKIKAVNPYGEAEANIK 4202
319 QY -----N-----N----- 320
4203 Db ITVIDKGPENITYPAVSRHTCTLNWDAPKDDGAEIAGYKIEYQEVSGSIQWKPGLI 4262
321 QY -----T-----F-----A-----S-----C-----ST----- 327
4263 Db SGTAYTVRGLEHQOQYFRFIRAEANAVGLSDYCOQVPVWIKDPDPFGCAPSTPEITGYDTN 4322
328 QY -----D-----M-----CI----- 331
4323 Db QVSLAMNPRDDGGSPILGYVVERFEKGGGWAPVKMPMVKGTECIVPGLHENETYQFR 4382
332 QY -----H-----V-----C----- 334
4383 Db VRVNAAGHGEPSNGSEPVTCRPYVEKPGADAPRVGKITKNSAELTWNRLRDLGGAPID 4442
335 QY -----KLK----- 337
4443 Db GYIVEKKLGDNDWTCNDKPYRDTAFVKNLGEKEEYFRVIAVNSAGEGEPSPDLV 4502
338 QY -----Q-----D-----R-----PI----- 342
4503 Db LIEEQRPIDFIDNLLKIDTVRAGETIQIRIPVAGGNPKPIIDLFGNSPIFENERTVVD 4562
343 QY -----K-----T-----F-----Q----- 346
4563 Db VNPGEIVITTTGSKRSDAGPYKISATNKYKDTCKLVNFVLDAPGKPTGPIRATDIQADA 4622
347 QY -----GH-----T-----N-----E-----V-----N 353
4623 Db MTLWRPXPKNCGDALNTYVVEKRTPGGDWTVVGHVPGVITLVRNLDANTPFEFRVAEN 4682
354 QY -----AI-----K-----W-----D-----P----- 359
4683 Db QYGVGEPLTDDAI VAKNPFDTPGAPQPEAVETSEAITLQWTRTPSDGGABIQGVVIE 4742
360 QY -----T-----GN----- 362
4743 Db KREVGSTWTKAAGFNILDTKHRVTGLTPKKTYPFRVAAYNAAGOGESYNSVPIITADNA 4802
363 QY -----L----- 363

Db	4803	PTRPKINMGMLTRDIIILAYAGERAKILVPPFAASGAPKVFPSKGENKISPTDPRVKVEYSDF	4862
Qy	364	-----L-----AS-----C-----	367
Db	4863	LATLTIEKSELTDGGLVFVELENSQGSASIRLKVVDPKASPQHIRVEDIAPDCCCTLYW	4922
Qy	368	---SD---DM-----T-372	
Db	4923	MPSSDGGSPITNYIIVEKDLRSHDGKWEKVSFVRNLNAYTVGGLIKONRYFRVRAETQ	4982
Qy	373	-----L-----K-----I-----W-----SMK-----Q-380	
Db	4983	YGVSEPCELADVUVAKYQFVPPNQPEAPTVRDKDSTWAELEMDPPRDGGS-KIIGYQVQY	5041
Qy	381	-D-----N-----CVDH---L-Q-----A-----389	
Db	5042	RDTSSEGRWINAKMDLSEQC-HARVTGLRQNGEFEPRIAKNAAGFSKPPSPPERCOLKSR	5100
Qy	390	-----H-----N-----K-----EI-Y-TIKW---S-----400	
Db	5101	FGPPGPPIHGAKSIGENHCTITWMAPLEDGGSKITGYNVVEIREYGSTL-WTVASDYNVR	5159
Qy	401	-P-T-----GP-----G-----T-----406	
Db	5160	EPEFTVKLREFNDYEPVVAINAAGKIGPSLPSGPIKIOESGSRPQIVVKPEDTAQPY	5219
Qy	407	N-----N-----P-----409	
Db	5220	NRRAVTCEAVGRPEPTARWLNRGRELPESSRYRFEASDGYVKTIKEVWDIDAGEYTV	5279
Qy	410	-N-----ANL-----413	
Db	5280	VSNPYSGDTATANLVQAPPVIEKDVNTILPSGDLVRLKIYFGSTAPFRHSILNAREE	5339
Qy	414	-M-----L-----A-S-----A-----S-----F-----420	
Db	5340	DMDHPTIRIVFDDHILITPALSVRAGRYEYTVSNDSGEATGFWLNVGTGLPEAPQGP	5399
Qy	421	-----D-S-----TV-----424	
Db	5400	LHISNIGPSTATLSWRPPVTDGSKITSYVVEKRDLSKDEWTVTNSVKOMNYIVTGLFE	5459
Qy	425	-----424	
Db	5460	NHEYFRVSAQENENGIGAPLVSEHPITARLPDPPTSPNLLEIVQVGDYVILSWQRPLS	5519
Qy	425	---RL-----W-----427	
Db	5520	DGGGLRGYIVEKQEBHEDFRCONPSPNNYVNPVLIDGRKYRVYFAVNDAGLSDL	5579
Qy	428	-D-----V-----D-----R-----431	
Db	5580	AELDOTLFAQSGGEGPKIVSPLSLNBEVGRCVTFECEISGSRPEYRWFKGCKELVDT	5639
Qy	432	-----G-----I-----C-----I-----H-436	
Db	5640	SKYTLINKDKQVLIINDLTSDDADEYTCRATNSGSTRANLRIKTKPRVFIPPKYVHG	5699
Qy	437	-----438	
Db	5700	GYESAQGETIELKIPYKAYPQGEARWTKDGEKIENNSKFSITDDKFAFLRISNASREDY	5759
Qy	439	-----438	
Db	5760	GEYRVVVENSGSDGTNNVTVADVPPEPRFPIENILDEAVILSWKPPALDGGSLVTNY	5819
Qy	439	-----T-----KH-Q-----E---PV---445	
Db	5820	TIEKREMGGSWSPKASRYTITTTIEGLRAGQYEFRIIAENKHGQSKPCFETAPVLIPG	5879
Qy	446	-----Y-----S-----V-----448	

RESULT, T 11

RESOL
S57242

S5/242 twitchin [similarity] - *Caenorhabditis elegans*

N-Alternate names: myosin-regulating protein

N;Alternate names: myosin-regulating protein kinase (EC 2.7.1.-)

QY 137 -----GA-H-----TI-A-N-----143
Db 2111 DQKIRIESEYKTRFVLGRATKRHAGLYTLTATNASGDKHSEVIVLKGPSPLGLEV 2170
QY 144 N-----H-----TDM-----ME-150
Db 2171 SNVYEDRADLEWKVPEDDGGAPIDHIEIEKMDLATGRWVPCGRSETTKTVPNLQGHY 2230
QY 151 -----V-----D-----G-----D-----154
Db 2231 KFRVRVANKGESDPLTTNTAILAKNPEYGVGRVDPKPELVDMKDHVDLAWNAPDDGGAP 2290
QY 155 VE-I-----P-----S-NKA-----162
Db 2291 IEAFVIEKKDKNGRWEALVVPDQKTATVPNLKEGEYQFRISARNKAGTGDSPSDR 2350
QY 163 VV-L-R-----G-----H-----ES-----170
Db 2351 VVAKPRNLAPRIHREDLSDTTVKVGATLKFIVHIDGEPADPTWSPNGKIGESKAQIEN 2410
QY 171 -----EV-----172
Db 2411 EPYISRALPKLRKQSGKYTITATNINGTDSVTINIKVSKPKPKPIEVTDPEDRA 2470
QY 173 -----P-----I-C-----175
Db 2471 TLDWKPPDDGGEFIEFIEKMKTKDGIWVPCGRSGDTHFTVDSLNGKDHYKFRKAVN 2530
QY 176 -----A-----176
Db 2531 SEGSPDLETETDILAKNPFDRDRGRPERPTDWDSDHVDLKWDPPLSDGAPIEEYQIE 2590
QY 177 -----W-----177
Db 2591 KRTKYGRWEPALTVPGGOTTATVPDLTPNEEYEFVRVAVNKGSPSDSASKAVIAKPRN 2650
QY 178 -----N-----PVS-----D-----L-----183
Db 2651 LKPHIDRALKNLTIKAGQISFDVPVSGEPAPVTWVHDPNREIRNGRVKLDNPEYQS 2710
QY 184 -LV-----SG-----S-GD-----190
Db 2711 KLVVKQMERGDSGFTTIKAVNANGEDATVKINVIDKPTSPNGPLDVSVDVHGDHVTLNR 2770
QY 191 -----S-----TA-----R-----194
Db 2771 APDDGGIPIENYVIEKYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVAANAE 2830
QY 195 -----I-----195
Db 2831 ESDPLEFTGTLKDPFKGTNAPEITDMDKHVDLEWKPPANDGGAPIEEYVEMKD 2890
QY 196 -----W-----NL-----S-E-----200
Db 2891 EFSFNDVAHVAGQTNATVGNLKEGSKYEFRIKAKNAGLGDPSDSASAVAKARNVPP 2950
QY 201 -----NS-----TSG-----PT-----Q-L-----V 210
Db 2951 VIDRNSIQEIKVAGQDFSLNIPVSGEPTTITWTFEGTPVESDDRMKLANNEDGKTRFH 3010
QY 211 -----LR-----H-----C-----214
Db 3011 KRALRSDTGTVIIKAENENGDTAEVKVTVLDPHSPSPGLDVTNIVKDGCDLAWKEPED 3070
QY 215 -----I-----215
Db 3071 DGGAEISHVIEKQDAATGRWTACGESKDTNFHVDLDTQGHEYKFRKAVNRHGDSDPLE 3130
QY 216 -RE-----217
Db 3131 AREALIAKDPFDRADKPGTPEIVDWDKDHADLKWTTPADGGAPIEGYLVEMRTSPGDWV 3190

QY 218 -----G-----218
Db 3191 PAVTVGAGELTATVDGLKPGQTYQFRVKALNKAQSPSPSRTWVAKPRHLAPKINRDM 3250
QY 219 -----QO-----DV-----P-----223
Db 3251 FVAQRVKAGQTLNFDVNVEGEPAPKIEWFLNGSPSSGNTNHNNTNNNTKLTTKSTAR 3310
QY 224 -----S-----N-----KQVT--S-----LD-----232
Db 3311 ADSGKYIVATNESKODHEVDVNILDIPGAPGEPRLRHKDIITKESVVLKWDPELDGGSF 3370
QY 233 -----W-----N-SE-----GT--L-----239
Db 3371 ITNYVVEKQEDGGRVWPCGETSDTSIKVNLSEGHYKFRKAVNRQGTSAPLTSDHAIV 3430
QY 240 -L-A-----T-----G-----SY-----D 246
Db 3431 AKNPPDEPDAPTDPVDMWDKDHVLEWKPPANDGGAPIDAYIVEKKDKFGDMVECARVD 3490
QY 247 G-----F-----A-----RI 251
Db 3491 GKTTKATADNLTPGETYQFRVKAVNKAQSPSPDPTGNVAVKPRMAPKLNLAGLLDLRI 3550
QY 252 -----W-----TKD-G-----NL-----A-----S-----T-L- 262
Db 3551 KAGTPIKLDIAFEGEPAPVAKWKANDATIDTGARADVTNTPTSSAIIHIFSAVRGDTGVYK 3610
QY 263 -----G-Q-----HK-G-----P-----I-- 269
Db 3611 IIVENHGXDTAQCNVTVLDPFTEGPIKIDIEIHKEGCTLAWKPPDNGGTDVLHYIVE 3670
QY 270 -----F-A-----L-----272
Db 3671 KMDTSRGTHQEVGTFPDCTAKVKNLVPGEYAFRVKAVNLQGESKLEAEPIIAKNQPD 3730
QY 273 -----KWN-----KK-----G-----278
Db 3731 VPDVDPKPEVDWDKDRIDIKWNPNTANNGGAPVTGYIVEKEKGSALWTEAGTPTGTTFS 3790
QY 279 --N-----F-----IL--S-----AG-----285
Db 3791 ADNLKPGYEFVRVIAVNAAGSPSDPTDPOITKARYLKPKILTASRKIKKAGTHNL 3850
QY 286 -----VD-K-TT-----290
Db 3851 EVDFIGAPDPTATWVDSGAALAPPELLVDKSSSTTSIFFPSAKRADSGNYKLKVNELG 3910
QY 291 -I-I-----W-----D-----294
Db 3911 EDEAIPEVIVQDRPSAPEGPLEVSDVTKDSVILNWKPPKDDGAEISNYVVEKRDTKNT 3970
QY 295 -----A-----H-----T-G-----E-AKQF-----304
Db 3971 WVPVSAFVTGTSITVPKLTEGHEYEFVRVMAENTFGRSDSLNTPDEVLAKDPFGTGPGR 4030
QY 305 -----P-----F-----306
Db 4031 PEIVTDNDHIDIKWDPDRDNGGSPVDHYDIERKDAKTGRWIKVNTSPVQGTAFSDTRVQ 4090
QY 307 --H-----SA-----P-A-----LD-----V-----314
Db 4091 KGHTEYRVVAVNKAQSPGSDSSAAATAKPMHEAPKFDLDLDGKEFRVKAGEPLVITIP 4150
QY 315 -----D-W-----Q-----S-----318
Db 4151 FTASQPDIDISWTEGKPLAGVETTSOTKLVIPTSTRSDSGPVKIKAVNPYGEAEANIK 4210
QY 319 -----N-----N-----320
Db 4211 ITVIDKEGAPENITYPAVSRHTCTTLNWDAPKDDGAEIAGYKIEYQEVGSQIWDKVPGLI 4270
QY 321 -T-----F-----A-S-C-----ST-----327

Db	4271	SGTAYTVRGLEHQYFRIRAEANAVGLSDYCOGQVWVWIKDPDPGAPSTBITGYDTN	4330	Db	5348	DMDHPTIRIVFDDHILITIPALSVREAGRYEYTVNSDGEATTGFWLNVTLGLPEAQGP	5407
QY	328	-----D-----M-----CT-----	331	QY	421	-----D-S-----TV-----	424
Db	4331	QVSLAWNPRDDGSPILGVYVERFEKRGGDWAPVPMVWKGTECIVPGLHENETYQFR	4390	Db	5408	LHISNIGPSTATLWSPVPTDGGSKITSYVVEKRDLSKDEWVTVTSNVKDMNYIVTGLFE	5467
QY	332	-----H-----V-C-----	334	QY	425	-----	424
Db	4391	VRAVNAAGHCEPSNGSEPVTCRPVVEKPGADAPRVGKITKNSAELTNWPLRDGGAPID	4450	Db	5468	NHEYEFVSAQNEENGIGAPLVSEBHPHIIARLPDPPTSPNLNLEIVQVGDVYVTLWSQRPIS	5527
QY	335	-----KLQ-----	337	QY	425	-----RL-----W-----	427
Db	4451	GYIVEKKLGDNDWTRCNDKPRDTAFEVKNLGEKEEYEFVAVNSAGEGEPKPSDLV	4510	Db	5528	DGGRLRGYIVEKQEBEHDEWFRCONQPNPNPNVFNLDGRKYRYRVFAVNDAGLSDL	5587
QY	338	-----Q-----D-----R-----PI-----	342	QY	428	-----D-----V-----D-----R-----	431
Db	4511	LIEQPGRPIDINNLKDTVRAGETIQIRIPVAGGNPKPIIDLFNGNSPIFENERTVVD	4570	Db	5588	AELDQTLFOASGSGEPKIVSPUSDLNEEVGRCVTFECEISGSPREYRWFKCKGKELVDT	5647
QY	343	-----K-T-----F-----Q-----	346	QY	432	-----G-----I-----C-----	436
Db	4571	VNPGIEVITTTGSKRSAGPYKISATNKYKDKCLNVFLDAPGKPTGPIRATDIOADA	4630	Db	5648	SKYTLINKGDKQVLIINDLTSDDADEVTCRATNSGSTRSTRANLRITKPRVFIPPKYHG	5707
QY	347	-----GH-----T-----N-E-V-----N	353	QY	437	-----TL-----	438
Db	4631	MTLSWRPPKNDGDAITNYVVEKRTPGGDWVTVGHPVGTTLRVNLDAANTPYEFVRVAEN	4690	Db	5708	GYEBAQGETIELKIPYKAYPQGEARWTKDGEKIENNSKFSITTTDDKFATLRISNASREDY	5767
QY	354	-----AI-K-----W-----D-----P-----	359	QY	439	-----	438
Db	4691	QYGVGPLETDDAIVAKNPDPFTPGAPQPEAVETSEEAITLQWTRPTSDGGAPIQGYVIE	4750	Db	5768	GEYRVVSVNSGSDSGTVNTVADVPPEPRFPPIENILDEAVILSWKPPALDGGSLVTNY	5827
QY	360	-----T-----GN-----AS-----C-----	362	QY	439	-----T-----KH-Q-----E-----PV-----	445
Db	4751	KREVGSTEWTKAAGNILDTHKRVTLGTPKKTVEFVAAVNAAGQGEYSVNSVPIADNA	4810	Db	5828	TIEKREMGSMSPCAKSRVYTTTISGLRAGKQYEFRIIENKHKQSKPCPEPTAPVLIPG	5887
QY	363	-----L-----	363	QY	446	-----Y-----S-----V-----	448
Db	4811	PTRPKINMGLTRDILAYAGERAKILVFAASPAKVTFSKGENKISPTDPRVKVEYSDF	4870	Db	5888	DERKREGRYDVBQGIKVRGKGTVSSNYDNYVPDIWKQYVQPEIKHDHVLHDYDHEE	5947
QY	364	-----L-----AS-----C-----	367	QY	449	-----AP-----S-----P-----	452
Db	4871	LATLTIEKSELTDGLYFVELENSQGSDSASIRLKVVDKSPASPOHIVERDIAPCCTLYW	4930	Db	5948	LGTGAFGVHVRTERATGNNAFAKFMVTPHESDKETVRKEIQTMSVLRHPTLNLHDAPE	6007
QY	368	-----SD-----DM-----I-----W-----SMK-----Q-----	380	QY	453	-----D-----G-----R-----Y-----	456
Db	4931	MPSSDGGSPITNYIVEKLDLRHSDGKWEKVSFVRNLNTVGLIKDNRYRFRVRAETQ	4990	Db	6008	DDNEMWMIYEFMSGGELFEKVADEHNKMSDEAVEYMRQVCKGLCHMHENNYVHLDLKE	6067
QY	373	-----L-----K-----I-----W-----SMK-----Q-----	380	QY	457	-----L-----A-----	458
Db	4991	YGVSEPELADVVAKYQFVPPNQPEAPTVRDKDSTWAELEWDPDRDGS-KIIGYOVQY	5049	Db	6068	NIMFTTKRSNELKLIIDFGLTAHLDPKQSVKVTGTGTAFAAPEVAEGKPVGYTDMMSVG	6127
QY	381	-----N-----CVHD-----L-Q-----A-----	389	QY	459	-----SG-S-F-----D-----K-C-----	465
Db	5050	RDTSGRWINAKMDLSEQ-HARVTLRQNGEPEFRRIIKNAAAGFSKPSPPSERCQLKSR	5108	Db	6128	LSYLLSGLSPFGENDETNRNVKSCDNMDDSAFSGISEDGKDFIRKULLADPNTRMT	6187
QY	390	-----H-----N-----K-----EI-Y-TIKW-----S-----	400	QY	466	VH-----I-----W-----N-----	470
Db	5109	FGPPGPIHVAGSIGRHNCTIITWAPLEDGSKITGYNVEIREYGSTL-WTVASDYNVR	5167	Db	6188	IHQALEHPLWTCNAPGRDSQIPSSRYTKIROSITKYDAWPEPLPLGLGRISNYSLLRKH	6247
QY	401	-----P-----T-----GP-----G-----	406	QY	471	-----TQ-----G-A-L-----V-H-----	478
Db	5168	EPEFTVDKLREFNDYFVRVAINAAGKIPSLPSGPIKIQESGSRPQIIVKPEDTAQPY	5227	Db	6248	RPOEYSIRDAFWDRSEAPRFIVKPYCTEVGEGQSANFYCRVIASSPPVVTWTKDRELK	6307
QY	407	-----N-----P-----	409	QY	479	-----S-Y-R-----G-T-----G-----G-----I-----PE-----	488
Db	5228	NRAVFTCEAVGPEPTARLWRNGRELPESSRYRFEASDGVYKFTIKEVMDIDAGSYTVE	5287	Db	6308	QSVKYMKNYNGNDYGLTINRVKDDKGEYTVRAKNSYGTKEIIVFLNVRHSHSPLKEPL	6367
QY	410	-----ANL-----	413	QY	489	-----V-----C-----	491
Db	5288	VSNPYGSDTANLVQAPPVIEKDVNTILPSGDLVRLKIYFSGTAPFRHSLVLRBEI	5347	Db	6368	EPKKKAPSPRVEEPKERSAPFTFHLNRLLIQNHQCKLTCSLOGNPNTIEWMKDGH	6427
QY	414	-----L-----A-S-----A-----S-----F-----	420	QY	492	-----NA-----AG-----D-----K-----	497
Db				Db	6428	PVDEDRVQVFRSGVCSLBIIFNARVDDAGTYTATNDLGDVDVSECVLTVQTKGSPIPR	6487

QY 498 V-----G-----A-----SA-----502
Db 6488 VSSFRPRAYDTLSTGTDVSHSYADMRRLSDVSDVRSAADLTKITNELPSFT 6547
QY 503 ---SD---GS-----V-----507
Db 6548 AQLSDSETVGGSAEFAAVSGOPEPLIEWLHNGERISEDSRFRASYVAGKATLRISDA 6607
QY 508 -----C-----VL---D-----LR--K 514
Db 6608 KKSDEGOYLCRASNSAGQEQTRATLTIVKGQDPLNGHAGQAVESELRVTK 6657

RESULT 12
T27935
hypothetical protein ZK617.1b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27935; T28031
R;White, S.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z20442
A;Accession: T27935
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-7160 <WIL>
A;Cross-references: UNIPROT:Q23551; EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617
A;Experimental source: clone ZK617
R;Harris, B.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z20458
A;Accession: T28031
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-7160 <WIL2>
A;Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b
A;Experimental source: clone ZK829
C;Genetics:
A;Gene: CESP:ZK617.1b
A;Map position: 4
A;Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 59
3067/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3
C;Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 70.5%; Score 2585; DB 2; Length 7160;
Best Local Similarity 8.0%; Pred. No. 5.9e-52;
Matches 470; Conservative 31; Mismatches 9; Indels 5360; Gaps 386;

QY 1 M-----S-----I-----S-S-----D-----6
Db 1113 MTEARRQSLFPGKKVEKWDIPLPKTVQQQVQKICEWKCTYSRPNAKIRWKDKRKEIFS 1172
QY 7 -----EV-----N-----FL-----11
Db 1173 GGLKXKIVIEKNVCTLIINNEVDVTGKYTCENGVPHTAQLTVLEPPMKYSFLNPLPNT 1232
QY 12 ---VVR---Y-----LOE-----S-G-----20
Db 1233 QEIVRTQAVLTCKVNTPRAPLVWYRGSKAIQEGDPRFIEKDVGRCITLTIKEVEBDQ 1292
QY 21 -----FS-----HSAF-----T-----27
Db 1293 AEWTARITQDVFSKVQVYVERPHT-FVVPKMSQKNESDLATLETLVNDKDAEVVWHD 1351
QY 28 -----F-----G--IESH-----33
Db 1352 GKRIDIDGVKFKVSSNRKRLLINGARIEDHGEYKCTTKDRTMAQLIVDAKNKFIVAL 1411
QY 34 -----IS-----QS-N-----I--N 40
Db 1412 KDEVIEKDDVTLMQTKDTKTPGIWFRNGKQISSMPGKPFETOSRNGHTHTLKGIEMN 1471

QY 41 -----G-----A-----L-----VP-----P 46
Db 1472 EADVYIEDOAGLRGSCNVTVLEAEKRPILNWKPKIEAKAGEPCVVKVPFQIKGTRGDP 1531
QY 47 -AA-L-----IS-----II-----Q-----54
Db 1532 KQAILKNGKPIDDEMRKLVVEVKKDDVAEIVFKNPQLADTGKWALELGSAGTALAPPEL 1591
QY 55 -----KG-LQ-----VV-----E-----A-----62
Db 1592 FVKDKPKPPKPLETKNTVTAEGLDLVWGTPDDEGAPVKAYIIEMOGRSGNWKVGETK 1651
QY 63 -----E-----V-----SI-----N-----67
Db 1652 GTDFKVKDLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPYGVGPKQNMDAIDVDKDH 1711
QY 68 -----EDG-----T-----L-----72
Db 1712 CTLAWEPPEDGGAPITGVYIERREKSEKDHQVGTQKCCCELTDKKYVEDKEYLYRVK 1771
QY 73 -----F-----D-----74
Db 1772 AVNKAGPGDCHGKPIKMKAKKASPEFTGGGKIKDLRLKVGETIKYDVPISGEPLPECLW 1831
QY 75 ---GRP-----IE-----S-----L-----S-----82
Db 1832 VVNGKPLKAVGRVMSSEKGMKIENAVRADSGKFTITLKNSSGSCDSTATVTVGPR 1891
QY 83 ---L-I---D-AVM---PD---VV---Q-----T 94
Db 1892 TPKGGLDIADVCADGATLSMNPDDGGDPLTGYIVEAQDMDNKGYIEVGKVDNNTTT 1951
QY 95 -----R-----O-Q-----97
Db 1952 LKVNGLRNKGYFRVKAVNNEGESEPLSADQVTQIKDPWDEFGKGRPEITDFDADRID 2011
QY 98 -----A---Y---RD-----101
Db 2012 IAWEPHKGGAPIEYIIVEVRDPTDKWKVKVRVDPDTNASISGLKEGKEYQFRVAVANK 2071
QY 102 ---K-LA-----104
Db 2072 AGPGQSEPEKQAKRPFIPAWLKHDNLKSIIVKAGATVRWEVKITGGPIEVEKVKGN 2131
QY 105 QQ-----H-----A-----A-----A-----110
Db 2132 QQLENGIQLTIDTRKNEHTILCIPSAMRSDVGEYRLTVKNSHGADSEKANKLTVLDRPSK 2191
QY 111 -----A-----A-----A-----A-----113
Db 2192 NGPLEVSDVPEDNLNLWSKPPDDGGEPYIYVEVEKLTATGRWVPCAKVKOTKAHIDGL 2251
QY 114 -----A-----AA-A-----T-----N-----119
Db 2252 KKGQTYQFRVKAVNKGASDALSTDKTKAKNPYDEPGKTGTPDVPDWDADRVSLEWEP 2311
QY 120 -----Q-----Q-G--S-----A-----K-N---G 127
Db 2312 KSDGGAPITQYVIEKKKGKGRDQWQEGKVSQDQTNAEILGLKEGEBEYQFRVAVNKAAGP 2371
QY 128 E-----N-----T-----AN-----GE-----E---N-----136
Db 2372 EASDPSRKVAKPNLKPWIDREAMKTTITIKVGNDFEFDVPVRGEPPEPPKKEWIFNEKFPVD 2431
QY 137 -----GA---H---TI-A-N-----143
Db 2432 DQKIRIESEYKTRFVLRGATRKHAGLYTLTATNAGSDKHSVEVIVLGPSSPLGLEP 2491
QY 144 N-----H-----TDM-----ME- 150
Db 2492 SNVYEDRADLEWKVPEDGGAPIDHVEIEKMDLATGRWVPCGRSETTKTTVPNLQPGHEY 2551
QY 151 ---V---D-----G-----D--- 154

Db 2552 KFRVAVNKEGSDPLTNTAILAKNPYEVGKVDKPELVWDKDHVDLAWNAPDDGGAP 2611
QY 155 VE--I--P--S--NKA--162
Db 2612 IEAFVIEKKDGRWEALVFGDQKATVPNLKEGEYQFRISARKAGTGDPSDSR 2671
QY 163 VW--L--R--G--H--ES--170
Db 2672 VVAKPRNLAPRIHREDLSDTTVKVGATLKPIVHIDGEPADVTWSPNGKIGESKAQIEN 2731
QY 171 ---F--I--C--EV--172
Db 2732 EPYISRALPKALRQSGKYTITATNINGDSVTINIKVSKPKPKPIEVTVDPEDRA 2791
QY 173 ---P--I--C--175
Db 2792 TLDWKPEDDGGPIEFYIEKANTKDWIWPGRSGDTHFTVDSLNGDHYKFRKAVN 2851
QY 176 ---A--176
Db 2852 SEGSPDLETTDILAKNPFDRPGRPEPTDWDSDHVDLKWDPPLSDCGAPIEYQIE 2911
QY 177 ---W--177
Db 2912 KRTYGRWEPAITVPGGQTTATVPDLTPNEEYFRVAVNKGSPSDASKAVIAKPRN 2971
QY 178 ---N--PVS--D--L--183
Db 2972 LKPHIDRALKNUTIKAGQISFDVPVSGEPAPVTVTHWPDNREIRNGRVLNDNPEYQS 3031
QY 184 -LV--SG--S--GD--190
Db 3032 KLVVKQMERGDSFTTIKAVNANGEDEATVKINVIDKPTSPNGPLDSDVHGHDVTLNWR 3091
QY 191 ---S--TA--R--194
Db 3092 APDDGGPIENYVIEKYDTASGRVPAKAVAGDKTTAVVDGLIPGHEYKFRVAANAEG 3151
QY 195 ---I--195
Db 3152 ESDPLETFTLAKDPFKGKNAPEITDWDKDHVDLEWKPPANDGGAPIEYVVMKD 3211
QY 196 ---W--NL--S--E--200
Db 3212 EFSFNDVAHVAGQTNATVGNLKEGSKYEFRIAKNAGLGDPSASAVAKARNVP 3271
QY 201 ---NS--TSG--PT--Q--L--V 210
Db 3272 VIDRNSIQEIKVAGQDFSLNIPVSGEPTTITWTPEGTPEVSDRMLNEDGKTFHV 3331
QY 211 ---LR--H--C--214
Db 3332 KRALRSDTGTYYIIKAENENGTDFAEVKVTVLDPSPRGLDVTNIVKDCDLAWKEPED 3391
QY 215 ---I--215
Db 3392 DGAIEISHYVIEKQDAATGEWTAGESKOTNFHVDLDTQGEHYKFRKAVNRHGDSDPLE 3451
QY 216 -RE--217
Db 3452 AREAIIAKDPFDRADKPGTPEIVDWDKDHADLKWTPPADGGAPIEGYLVMETPSGDWV 3511
QY 218 ---G--218
Db 3512 PAVTVGAGELTATVDGLKPGQTYQFRKALNKAGESTPSPSRMTVMKPRHLAPKINRDM 3571
QY 219 ---GO--DV--P--223
Db 3572 FVAQRVAKAGTGLNFDVNVEGEPAPKIEWFLNGSLSSGGNTHIDNNTKLTTKSTAR 3631
QY 224 ---S--N--KDV--S--LD--232

Db 3632 ADSGKYKIVATNESGKDEHEVDVNIILDIPGAPEGPLRHKIDITKESVVLKWDPEPLDDGGSP 3691
QY 233 ---W--N--SE--GT--L--239
Db 3692 ITNYVVEKQEDGRWVPCGSETSDTSUKVNLSEGHYKFRKAVNRQSTAPLTSDBAIV 3751
QY 240 ---L--A--T--G--SY--D 246
Db 3752 AKNPPDEPADTDTVPDWDKDHVDLEWKPANDGGAPIDAYIVEKKDFGDWVECARVD 3811
QY 247 G--F--A--A--RI 251
Db 3812 GKTTKATADNLTPGETYQFRKAVNKGKPSDPTGNVAVKPRMAPKLNLAGLIDLRI 3871
QY 252 ---W--TKD--G--NL--A--S--T--L--262
Db 3872 KAGTPIKLDIAPEGEPAVPAKWKANDATIDTGARADVTTPTSSAHIHFSAVRGDTGVYK 3931
QY 263 ---G--Q--HK--G--P--I--269
Db 3932 IIVENEHGKDTAQCNVTVDLPCTPEGLKIDIEIHKEGCTLNKPKPTDNGGTDVLHVIVE 3991
QY 270 ---P--A--L--272
Db 3992 KMDTSRGTVQEVGTPDCTAKVNLKVPGEYAFRKAVALQGESKPLEAEPPIAKNQFD 4051
QY 273 ---KWN--KK--G--278
Db 4052 VPDVDKPEVTDWDDKDRIDIKNPTANNGAPVTVGIIVEKKEGSAIWEAGTGTPTFS 4111
QY 279 ---N--F--IL--S--AG--285
Db 4112 ADNLKPGVEYFRVIAVNAAGSPDSDPTDPOITKARYLKPILKTASRKIKIKAGFTHNL 4171
QY 286 ---VD--K--TT--290
Db 4172 EVDFIGADPTATVTVGDSGAALAPELLVDKSTSTSIFFPSAKRADSGNYKLVKNELG 4231
QY 291 ---I--I--W--D--294
Db 4232 EDEAIEVIVQDRPSAPEGLEVDVTKDSCVLNWKPPKDDGGAIESNYVVEKRDYKNT 4291
QY 295 ---A--H--T--G--E--AKQF--304
Db 4292 WVPVSFAVTGTSITVPKLTGEGHEYFRVMAENTFGRSDSLNTDPEVLAKDPFGTGGKGR 4351
QY 305 ---P--F--306
Db 4352 PEIVDTDNHDHIDIKWDPPRDNGSPVDHYDIERKDAKTGRWIKVNTSPVQGTAFSDTRVQ 4411
QY 307 ---H--SA--P--A--LD--V--314
Db 4412 KGHTEYRVVAVNKGAPGQPSDSSAAATAPKMPHEAPKFDLDDGKEFRVKGAGPLVITIP 4471
QY 315 ---D--W--Q--S--318
Db 4472 FTASQPDLSWKEGKPLAGVETDTSQTLVLPSTRRSDSGPVKIKAVNPYGEAEANIK 4531
QY 319 ---N--N--320
Db 4532 ITVIDKPGAPENITYPAVSRHTCTLNWDAPKDDGGAIEAGYKIEYQEVGSIQWDRVPLGI 4591
QY 321 ---T--F--A--S--C--ST--327
Db 4592 SGTATVIRGLEHQYFRIRAEAVGLSDYCOGVPVWIKDPPDPGAPSTPEITGYDTN 4651
QY 328 ---D--M--CI--331
Db 4652 QVSLAWNPPDDGSPILGYVVERFEKGGDWAPVMPMPVKGTCEIVPGLHENETYQPR 4711
QY 332 ---H--V--C--334
Db 4712 VRVNAAGHGPESNGSEPVTCRPPYVEKPGADAPRVGKITKNSAELTNWRLRDGGAPID 4771

QY 335 -----KLG----- 337
Db 4772 GYIVERKLGNDWTRCNDKPRDPTAFVKNLGEKBEYEFVRVAVNSAGEGEPKSDLV 4831
QY 338 -----Q-----D-----R-----PI----- 342
Db 4832 LIBEQGRPIFDINNLKDITVRAGETIQIRIPVAGGNPKFIIDLFGNNGPIFENERTVVD 4891
QY 343 -----K-T-----F-----Q----- 346
Db 4892 VNPGEIVITTTGSKRSAGPYKISATNKYKDKTCKLVNFDAPGKPTGPIRATDIQADA 4951
QY 347 -----GH-----T-----N-B-V-N 353
Db 4952 MTLSSRPKPKNGGDAITNVYVEKRTGGDWTVGHPVGTTLRVRLNDANTPYEVRVAEN 5011
QY 354 -----AI-K-----W-----D-P----- 359
Db 5012 QYGVGPLETDDAIVAKNPPDTPGAPGQPEAVETSEAITLQWTRPTSDGAPIQGYVIE 5071
QY 360 -----T-----GN----- 362
Db 5072 KREVGSTEWTKAAGNILDTKHRVTGLTPKTYEFRVAAYNAAGQGEYSVNSVPITADNA 5131
QY 363 -----L----- 363
Db 5132 PTRPKINMGMLTRDILAYAGERAKILVPPAASPAPKVTFSKGENKISPTDPRVKVYSDF 5191
QY 364 -----L-----AS-----C----- 367
Db 5192 LATITIEKSELDTGGLYFVLENSQGSASIRLKVVDPKASPQHIRVEDIAPDCCTLYW 5251
QY 368 -----SD-----DM----- 372
Db 5252 MPSSDGGSPITNVIVEKDLRHSQKWEKVSFVRNLNTVGLIKDNRYRFRVRAETQ 5311
QY 373 -----L-----K-----I-----W-----SMK-----Q----- 380
Db 5312 YGVSEPCELADVVAKYQFEVNPQPEAPTVRDKDSTWAELEWPPRDGGS-KIIGYQVQY 5370
QY 381 -D-----N-----CVHD-----L-O-----A----- 389
Db 5371 RDTSSGRWINAKMDLSQC-HARVTGLRQNGEPEFRIIAKNAAGFKSPSPSERCQLKSR 5429
QY 390 -----H-----N-----K-----EI-Y-TIKW-----S----- 400
Db 5430 FGPPGPIHVGAKSIGRNHCTITWMAPLEDGGSKIYGNVEIREYGSTL-WTVASDYNVR 5488
QY 401 -P-T-----GP-----G-----T----- 406
Db 5489 EPEFTVDKLRFNDEYFRVAINAAGKIPSLPSGPIKIQESGGRPQIVVKPEDTAQPY 5548
QY 407 N-----N-----P----- 409
Db 5549 NRRAVTCEAVGREPTARMLNRGRELPESSRYRFEASDGVYKTIKEVWDIDAGEYVE 5608
QY 410 -N-----ANL----- 413
Db 5609 VSNPYGSDTATANLVQAPPVIEKDVNTILPSGLVRLKIYFGTAPFRHSLVLRNEEI 5668
QY 414 -M-----L-----A-S-----A-----S-----F----- 420
Db 5669 DMDHPTIRIVEFDHILITIPALSVREAGRYEYTVSNDSEBATTGFWLNVTLGULPEAPQGP 5728
QY 421 -----D-S-----TV----- 424
Db 5729 LHSNIGPSTALSWRPVTDGSKITSYVVEKEDLSKDEWTVTVSNVKDMNYIVTGLFE 5788
QY 425 ----- 424
Db 5789 NHEYEFVSAQNGIGAPLVSEHPHIIARLPDPPTSPNLNLEIVQVGVDYVTLSSWRPLS 5848

QY 425 -----EL-----W----- 427
Db 5849 DOGGRLRGIYVEKQBEEDHEWFRQNFQNPNNVFNLI DGRKYRYRFAVNDAGLSDL 5908
QY 428 -----D-----V-----D-----R----- 431
Db 5909 AELDTLFOAGSGGPKIVSPLSLDNEBVGCVTFECEISGSPREYRWFPGCKELVDT 5968
QY 432 -----G-----I-----C-----I-----H----- 436
Db 5969 SKYTLINKGDKQVLIINDLTSDDADEYTCRATNSGSTRANLRITKTRVFI PPKYHG. 6028
QY 437 -----TL----- 438
Db 6029 GYBAQGETIELKIPYKAYPQGEARWKDGEKIENNSKPSITDDKFAFLRISNASREDY 6088
QY 439 ----- 438
Db 6089 GEYRVVVENSGSDGTNVNVTADVPEPRFPIENILDEAVILSMKPPALDGGSLVTNY 6148
QY 439 -----T-----KH-Q-----E-PV----- 445
Db 6149 TIEKREAMGGSWSPCAKSRYYTTTIEGLRAGQYEFRIIAENKHGOSKCPCEPTAPVLI PG 6208
QY 446 -----Y-----S-----V----- 448
Db 6209 DERKRRGYDVDEQGIKVRGKGTSSNYDNYVDIWKQYYPQVPEIKHDHLDHYDIHEE 6268
QY 449 -----AF-----S-----P----- 452
Db 6269 LGTGAPGVHVRTERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAPE 6328
QY 453 -D-----G-----R-----Y----- 456
Db 6329 DDNMVMIYEFMSGGELFEKVADEHNKMSDEBAVEYMRQVCKGLCHMHNYYVHLDLKEE 6388
QY 457 -----L-----A----- 458
Db 6389 NIMFTTKRSNELKLIDFGLTAHLDPKQSVKVTGTGTAFAAFAEVAEGKPVGYTDMWSGV 6448
QY 459 -----SG-S-F-----D-----K-C----- 465
Db 6449 LSYILLSGLSPFGGENDETLRNVKSCDWNMDSAFSGISEDGKDFIRKLLADPNTRMT 6508
QY 466 VH-----I-----W-----N----- 470
Db 6509 IHOALEHPMLTPOGNAFGRDSQIPSSRYTKIRDSIKTKYDAWPEPLPLGLRISNYSRLKH 6568
QY 471 -----TQ-----T-----G-A-L-----V-H----- 478
Db 6569 RPOEYSIRDAFWDRSEAPRFIVKPYGTEVGEQSANFYCRVIASSPPVVTWHKDDRELK 6628
QY 479 -S-Y-R-----G-T-----G-----G-I-----FE----- 488
Db 6629 QSVKMKRYNGNDYGLTINRVKGGDKGEYTVRAKNSYGTKEEIVFLNTRHSEPLKFEPL 6688
QY 489 -----V-----C-----W----- 491
Db 6689 EPMKAPSPRVEEFKERRSAPFTTHLRNRLIQKNHQCKLTCSLQGNPNPTIEWMKDGH 6748
QY 492 -----NA-----AG-----D-----K----- 497
Db 6749 PVDEDRVQVSFRSGVCSLEIFNARVDDAGTYTATNDLGDVDSVCEVLTVOTKGPPIPR 6808
QY 498 V-----G-----A-----SA----- 502
Db 6809 VSSFRPRAYDTLSTGTVDERSHAYDMRRRLSRIRDVSPDVSRAADDLTKITINELPSFT 6868
QY 503 ---SD-----GS-----V----- 507
Db 6869 AQLSDSETVGGSAEFAAVSGOPEBLIEWHNGERISBDSDFRASYAGKATLRISDA 6928
QY 508 -----C-----VL-----D-----LR-K 514

Db 6929 KXDEGQYLCRASNSAQEQTRATLVKGQPLNGHAGQVSELRVTK 6978
RESULT 13
T42737
gp330 protein precursor - rat
N:Alternate names: megalin
C:Species: Rattus norvegicus (Norway rat)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42737
R:Saito, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A>Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of the gp330 family
A:Reference number: A58173; MUID:95024033; PMID:7937880
A:Accession: T42737
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4660 <SAI>
A:Cross-references: UNIPROT:P98150; EMBL:L34049; NID:G561852; PID:G561853; PIDN:AAA51369
A:Experimental source: strain Sprague-Dawley; kidney
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
F:1-25/Domain: signal sequence #status predicted -SIG-
F:26-4660/Product: gp330 protein #status predicted <MAT>
Query Match 70.4%; Score 2583.7; DB 2; Length 4660;
Best Local Similarity 9.8%; Pred. No. 2.7e-52;
Matches 428; Conservative 64; Mismatches 13; Indels 3843; Gaps 370;
Db 1 M-S-I-S-SDE-V-8
113 MTCNSGQICPSEYRCHDVSDCPDGSBERNCHYPTCDLTGANGACYNTSQRCDQKVD CRD 172
QY 9 -N-P-L-VY-R-Y-15
Db 173 SDEACTLCSQKEFCGEGCEILRAYVCHDNDCEDESDERNVYDTCGHQFTCSNG 232
QY 16 -L-Q-ES-19
Db 233 QCINQNVCDGDDCQDSGDEDCESNQSHRCYPREWACPGSGRCISIDKVCVGVPDCP 292
QY 20 -G-F-S-22
Db 293 EGDENNVTSGRTCGMGVCSVLNCEYCHOTPFEGCEFCPPGHIINSNDSRTCIDFDDCQ 352
QY 23 -H-S-AFT-P-29
Db 353 IWGICQKCNROGRHQCLCEBGIYLERGHQCKSSDSFSAASVIFNGRDLVGLHGRN 412
QY 30 -I-ES-H-I-SQ-S-N-38
Db 413 FRILAESKNRGWVGDFHVKHRVFTDPMQEKVFTDINGLNTQELNVSDVTPENLA 472
QY 39 -IN-AL-V-44
Db 473 VDWINKLVETKVNRI DVNLEGNQVTLITENLGHPRGIALDPTGVLYFPFSDWGSLS 532
QY 45 -P-PA-ALIS-I-IQ-K-55
Db 533 GQPKVERAFMDGNRKDLVTKVGVNAGITLDLVSKRVVWDSRYDIETVTVDIGRKT 592
QY 56 -G-L-FEESHVFTDVKAVMKASKFTETNPQVTHQSSLRPHGVTVYHA 59
Db 593 VARGSLVPHFGISLFEESHVFTDVKAVMKASKFTETNPQVTHQSSLRPHGVTVYHA 652
QY 60 -V-E-AE-VSI-N-67
Db 653 LRQPNATNPGCSNNGCAQCVLSHRTDNGGLGYRCKEFGFELDDDEHRCVAVKFLLF 712
QY 68 -E-1
Db 713 SSKTAVRGIPFTLSTQEDVMVPTGSPFPFVGIDFDAQHSTVFYSDLSKDIYKQKIDGT 772

QY 72 -L-FD-74
Db 773 GKEVITANRLSEVECLTFDWISRLNYWTDGGLKSVTVLRLADKSRRIISLNNRPSIVV 832
QY 75 -G-PP-I-E-S-L-SL-I-DA-86
Db 833 HPTAGYMFSLSDWFRPAKIMRANSDGSHLMFIVNTSLGWPNGLAIDMSASRLYWDADFDDK 892
QY 87 -V-M-P-D-V-VQ-T-R-95
Db 893 IEHSTILDGLDRKELGHVDQWTHPFGITVFKDNVFIIDWRLGAIIRVNRKSDGDMVTIIRRG 952
QY 96 -Q-Q-A-Y-RDKL-103
Db 953 ISSVMHVKAYDADLQTSNYCSQTHANGDCSHFCFPVPNFQVRCGCPYGMKLQDQWTC 1012
QY 104 -A-QQ-H-A-A 109
Db 1013 EGDPAEPPTQCGSLSPFNNGKCVPSFRCDGVDDCHDNDCEHOCGVFNNTCPSAFA 1072
QY 110 -A-110
Db 1073 CVRGGQCIPOGWHCDRQNDCLDGSDEQNCPTHTATSTCTPSTCDNHVCIPIKDWVCDTD 1132
QY 111 -A-A-A-AAA 116
Db 1133 NDCSGSDEKNCQASGTQCPQFCDHRCISPLVYVCDGDKDCADGSDGAGCVLNTCSAQ 1192
QY 117 -A-T-N-Q-Q-121
Db 1193 FKADGSSCINSRYRCDGVYDCRNSDEAGCPTPPGMCHPDEFQCGDGTCTIPNTWEC 1252
QY 122 -GS-A-K-NG-E-128
Db 1253 GHPDCIHGDEHTGCVPTKCSPTHFLCDNGNCIYKAWICDGNDCRMSDEKDCPTQPFH 1312
QY 129 -N-TA-NG-E-E-AH-139
Db 1313 CPSTQWQCPGYSTCINLSALCDGVFCPNGTDESPLCNQDCSHFNGGCTHQCQWGPFGA 1372
QY 140 -T-I-A-142
Db 1373 TCLCPGLYQLANDTKTCEDINECDIFGCSQHCVMNRGSRFCADPEYTLSDGRTCKVT 1432
QY 143 -N-N-H-145
Db 1433 GSENPILLVVASRDKIIVDNITAHNTNLSLVQDVSVFVALDFDSVTGRVFWSDLLQGTW 1492
QY 146 -TDM-M-EV-DG-153
Db 1493 SVFQNGTDKRVVHDSGLSVTEMIADVWIGRNLYWTDYALETIEVSKIDGSHRTVLISKV 1552
QY 154 -D-V-E-I-P-S-N-160
Db 1553 TKPRGIALDPRMGDNVFWSDMGHHPRIERASMDGTMTRTVIVQEKIYWPCLSIDYPNRL 1612
QY 161 -K-A-V-VL-RG-167
Db 1613 IYFMDAYDIIEFCDYDGHNRQVIAASDLVHLHPHALTLFEDFVYVWTDRTQVMOQANKW 1672
QY 168 H-ESEV-F-I-CA-176
Db 1673 HGGNQS-VVMYSVHQPLGITAHPSPQPSRNPNCASCSCHLCLLSAQAPRHVSCAPSG 1731
QY 177 WN-PVS-D-182
Db 1732 WNLSDSVNCRGDOPFLMSVRDNIIFGISLQPEVKSNDAMVPSIGIHYGVDFDDEQ 1791
QY 183 -LL-V-S-GS-G-D-S-T-A-193
Db 1792 FIYVWENPGEIHRVKTGDSNRVTFAPLSLLGSSLSGLALDQVSRNIYYTTTPASRSEIVLTL 1851
QY 194 -R-I-W-NL-S-199

Db	1852	KGTRVKGKTLIANDGTPLGVGPPVGLAVDPARGKLYWSDHGTDSGVPKAKIASANMDGTSI	1911
Qy	200	E-----N-----S-----TS-----G-----P-----	206
Db	1912	KILFTGNLQHLVVLTDIOEQKLYAWTSRGVIERGVNDGTERMIHLVHHLAHPGLWVYG	1971
Qy	207	-----T-----Q-----VLR-----H-----C-----	214
Db	1972	SFLYSDQEVIERVDKSGNNKVVLRDNVPLRGLRVHRRNAADSSNGCSNNPACQ	2031
Qy	215	I-----R-----E-----G-----	218
Db	2032	QICLPVPGMFCACASGFKLSPDGRSCSPYNSFMVMSLPAVRGFSLELSHDSEAMVPV	2091
Qy	219	-GO-----DV-----P-----SN-----K-----	226
Db	2092	AGQGRNVLHADVDVANGFIWCDPSSSVRSNGIRRIKPDGSNFTNVVTVYGANGIRGV	2151
Qy	227	-D-----V-----TS-----L-----D-----W-----	233
Db	2152	ALDWAAGNLYFTNAFYETLIEVLIRINTTYRRVLLKVSVDMPRHIIIVDPKHYLFWADYG	2211
Qy	234	-----N-----SEG-----T-----LLA-----TGSY-----DG-----	247
Db	2212	QPKIERSFLDCTNRTVLVSEGIVTPRGL-AMHDGTG-YIYWVDDSLDLIARIHLDGGS	2269
Qy	248	-----F-----AR-----TWKD-----GN-----	257
Db	2270	QVRYGSRYPYPTGIVTFGESIIWV-DRNLKVKFOASKQPGNTDPPVIRDKINLRDVT	2328
Qy	258	-----L-----A-----S-----TLG-----O-----	264
Db	2329	IFDEHAQPLSPAELNNPCLQSGGCHFCFALPELPTPRCGCAFTLGNHGKSCATSQE	2388
Qy	265	-----H-----K-----	266
Db	2389	DFLIYSLNLSLHFPDRHSLPQVIVSAGTALDLYDRNNRIFFFTQKLSLRQIS	2448
Qy	267	-----G-----P-----I-----F-----	270
Db	2449	YVSLYSGSSPTVLLSINIGVTDGIAFDWINRRIYSDFSQNTINMAEDGSNRAVIARVS	2508
Qy	271	-----A-----LK-----W-----N-----K-----K-----GNF-----IL-----S-----AG-----	285
Db	2509	KPRAIVLDPCRGYMTWGTNNAKIERATLGGNFRFVIVTSLVWPNGLALDLETDLIYW	2568
Qy	286	VD-----K-----T-----TI-----I-----W-----D-----	294
Db	2569	ADASLQIERSTLTGTNRVVVSTAFHSFGLTVYGVQIYVTDLYTRKIYRANKYDGSILV	2628
Qy	295	A-----H-----TG-----E-----A-----	300
Db	2629	AMTTRLPQSGISTVVKTORQOCNPDOPNGGCHICAPGNPARGCQCPHEGNWYLAN	2688
Qy	301	-----K-----O-----OF-----P-----PH-----S-----A-----	309
Db	2689	DNKYCVVDTGRCNLOQFTCLNGHCINQDWKCDNDNDGCGSDELPTVCAFTCRSTFT	2748
Qy	310	-----P-----A-----L-----	312
Db	2749	CGNRCVPYHYRCDYVNDGDNSEAGCLPNCNMTTEFTCSNGRCIPLSVYVCGNINCH	2808
Qy	313	-----D-----V-----D-----	315
Db	2809	DNDTSDEKNCPPTCPDFTKQCTTNCVPRAFLCDGNDGCGSDENPIYCASHTCRSN	2868
Qy	316	-----W-----O-----S-----	318
Db	2869	BFQCLSPORCIPSWFCDEADGADGDEPDTGHSVNTCRASQFQCDNGRCISGNWVCD	2928
Qy	319	-----N-----T-----F-----	322
Db	2929	GDNDGDMDSDEQRHHCELQNCSTQFTCVNSRPPNRRICIPOYWVCDGDADCDALDELQ	2988
Qy	323	-----A-----SC-----	325
Db	2989	NCTMTCSAGFSCANGRCVROSFCDRRNDGCDYSDERGCSYPPCHANQFTCQNGRCIP	3048
Qy	326	-----S-----T-----D-----MCI-----HV-----C-----	334
Db	3049	REFVCDNDGDSDEQEHLCHTPEPTCLPHQFRCDNGHCIEMGRVCNVHVDGDSNDE	3108
Qy	335	K-----L-----G-----Q-----	338
Db	3109	KGGINECLDSSISCDHNCTDITISFYCSCLPFGKLMDSKSCVDIDECKESPQLCSOK	3168
Qy	339	-----D-----R-----P-----I-----	342
Db	3169	CENVGSYICKAPGYIREPDGSKCRQNSIBPYLIFSNRYIRNLTDTGSSYLILOGL	3228
Qy	343	-----KT-----	344
Db	3229	GNVVALDFRVEKRLYWIDAQKIIIRMFNKTNRETIINHRLRAESLAVDWVSRKLYW	3288
Qy	345	-----F-----Q-----GH-----	348
Db	3289	LDAILDCLFVSDLEGRHRRKMIQAHCVDANNTFCFEHPRGIVLHPQRGHVYVADWGVHAYI	3348
Qy	349	-----TNE-----V-----NAI-----KW-----D-----	358
Db	3349	GRIMDGTNKSIIITKIEWPNAITIDYNDLLYWAHAHLGYIEFSDLEGHHRTVYDGS	3408
Qy	359	-----P-----T-----GN-----L-----LAS	366
Db	3409	LHPFALTIFEDTVFTWDNTRTVKGNKYDGSGRVLVNTTHKPPDIHVHYHPYQPIMS	3468
Qy	367	-----CS-----DD-----M-----T-----L-----K-----	374
Db	3469	NFCGTNNGCSHLCLIKAGRGFTACPDQFOTVQLRDLTCLMPMCSQFCLGNNKCI	3528
Qy	375	-IW-----S-----MK-----Q-----D-----NC-----	383
Db	3529	PIWKCQDQKDCSDGSDPDLCPHRCFLGQFCQCRDGNCTSPQALCNARQDCADGSEDR	3588
Qy	384	V-----H-----D-----L-----Q-----AH-----N-----	391
Db	3589	VLCHEHRCESNEWQCANCKICIPQSWQDCSVNDCLNSDEDTSHCASRTCRPGQFKCNGR	3648
Qy	392	-----K-----E-----I-----Y-----T-----IK-----W-----	399
Db	3649	CIPQSWKCDVNDGCDYSDEPIDECTTAAAYNCNDHTEFCKTNYRCIPOMAVCNGFDDCR	3708
Qy	400	-----S-----PTG-----P-----GT-----N-----N-----	408
Db	3709	DNSDEGCSVPCHPSGDFRCAHHCIPLRWKCDGTDGCDNSDEENCVPRECSSEFRC	3768
Qy	409	-----P-----N-----AN-----LM-----LA-----SA	418
Db	3769	ADQCIPSRWVCDQDNDGSDERDCENKTCHEPHFQCTSGHCVKALACDGRADCLDA	3828
Qy	419	SFD-S-----TVR-----LW-----D-----VDRG-----I-----C-----I-----	435
Db	3829	S-DESACPT-RFPNGYTCPAAMPECKNHVICOSFIWCDGENDCVD-GSDEEHLCLFNPIC	3885
Qy	436	-----HTL-----T-----K-----H-----QBPV-----Y-----S-----	447
Db	3886	ESPQFRCDNSRCVYHQLCNGVDDCGDGSDEKEEHCRTKPKCTDTTEYKCSNGNCISQ	3945
Qy	448	-----V-----A-----F-----S-----P-----	452
Db	3946	HYVCDNVNDCGLDSTGCLNLDNRCAENICEQNTQLSGGGFCISCRPGFKPSTSDKN	4005
Qy	453	-----D-----G-----VYLAS-----GS-----P-----D-----K-----C-----	465
Db	4006	SCODINECEBFGICPOSCR--NSKSGYECFCVDGFKSMSTHYGERCAADGSPFLLLPEN	4063

QY 466 V-----HI-----WN-----T-Q-T-GA-----475
Db 4064 VRIRKNTSSEKFELEEHIOTIDYDNDPEHIGLSVVYVTVLAQSGFAGKRAYIP 4123
QY 476 -----L-----V-H-----S-----YR-----481
Db 4124 NFESGNNPIREVDLGLKLMQPDGLAVDWGRHIYWSDAKSORIEVATLDGRYRWLIT 4183
QY 482 -----G-----T-----G-----G-----484
Db 4184 TQLDOPAAIAPNPKLGLMFTDQKQPKIESAMWNGEHRVLSNGLWPNGLSIDVND 4243
QY 485 -----G-----G-----IFE-----V-----C-W--NAAG 495
Db 4244 DRVWSDSKEDVIEAIKVDGTDRLRIINEAMKPFSLDIPEDKLYWVAKEGEVWRQKFG 4303
QY 496 ----DKV-----LD-----GA-S-A-----SD-----504
Db 4304 KENKEVLVVPMLTOVRIFHQLRYNSQSVNPKQVCSHLCLLRPGGYSACPGSDFVT 4363
QY 505 GS-V-C-----V-----LD-----511
Db 4364 GSTVQCDAASELPTVPPPCRMHGGNCYFDENELPKCKCSSGYSGEYCEVGLSRGIPPG 4423
QY 512 -----LTK 514
Db 4424 TTMVALLTFVITVIGALVLVGLFHYRK 4451

RESULT 14
T12117
polyprotein - fava bean dsRNA replicon
C:Species: Vicia faba (fava bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12117
J. Gen. Virol. 79, 2349-2358, 1998
A:Title: Nucleotide sequence, genetic organization and expression strategy of the double
A:Reference number: 217424; MUID: 98451319; PMID: 9780039
A:Accession: T12117
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5825 <PFE>
A:Cross-references: UNIPROT:O82731; EMBL:AJ000929; NID:g3184155; PIDN:CAA04392.1; PID:g3
A:Experimental source: virion; cultivar 447
C:Comment: This gene product may be cleaved into several proteins including helicase and
C:Genetics:
A:Genome: dsRNA replicon
C:Superfamily: fava bean dsRNA replicon polyprotein

Query Match 70.3%; Score 2580.1; DB 2; Length 5825;
Best Local Similarity 9.1%; Pred. No. 5.2e-52;
Matches 442; Conservative 56; Mismatches 13; Indels 4349; Gaps 378;
QY 1 M-----SIS-----S-----D-----E-----V-----N-- 9
Db 949 MATPESITQAPVKESNTLLIDMNLMDPMWLIECLRRKITAYVTPMLELQNPD 1008
QY 10 -----F-----LV-----Y-----13
Db 1009 QTILGDEVQFRKENQIVITAWSRPIRVTTENYMNWTKVDPAFVIGENTPIDTVREAN 1068
QY 14 -----R-----Y-----13
Db 1069 GISLRIDLAASSTFEMSEVETVGSIEISVPTVYVYGGPKRKLMSGEEKVKLHREL 1128
QY 19 -----SG-----FS-----HS-AF-----26
Db 1129 YRRMIRNVSGTSPFKELVDYIGISNMKNINDMVVSFSDISDDTIRHTAATVYSKMIT 1188
QY 27 -----T-----F-----28

Db 1189 RANSMOLTERVLESVNTARGFLASLLISOAETMLKITGLDETMLTTVNRWMSKSGSEMVRL 1248
QY 29 -----G-----G-----I-----E--- 31
Db 1249 KGLMNMINSLDIDNLMCYSRPVIHRLRAOETLEPTGTCPHKTPDFVKECSMCSCCG 1308
QY 32 -----S-----H-----33
Db 1309 VNPAIGAIGFSSCLPTDKHQCDHPCKCHAHSEAGTRCSCCLPIAGEACPCCGVNRQIE 1368
QY 34 --I-----S-----Q-----SNIN--- 40
Db 1369 SEILFENSEADEAEQNRVRKNSKPRPRDREKGNRNNRNDSSRRRTDANHHNSVNYHH 1428
QY 41 -----GALV--PP--AAL-----I-----S-----I 52
Db 1429 GHNKPORQGA--TOOPPRSAALHPENDNDTPDSTPTVILTADPTTNPNGEROAPGDNI 1487
QY 53 -----I-----Q-----KG-----L--Q-----Y---V 60
Db 1488 PHENDIPGSSSTTTQSSPPDDTNYSGPEHRNRIQNHRIIAQTNREFFGLDYAYGNDIAV 1547
QY 61 -----E-----A-----EV-----S-----I---N 67
Db 1548 KKNSHKLMPPNEQRLRSCFQFMNMGDSIYATENLEVIKTYNTSGEGYCGYNALKILYPN 1607
QY 68 -----E-----D-----69
Db 1608 LDLTLEEMQEI VGSQTFQSDWEIMRVAQAKQLNLI VTERCALVNSKVCNEFGVICHCR 1667
QY 70 --G-----TL-----F-----D-----G-----R-----P--- 77
Db 1668 HRGVMLEHWEAALAIQKPADYHPTFTNALTREDLLDFAKSSGLNKNKVRFSVLGDPRL 1727
QY 78 -I-----ESLS-LI-D--A--VM-----P-----D-----D--- 90
Db 1728 KLQTEMHESLSKVVEDVNPAGFVKITKGSTHYVTNDRNLGYSPOYGTFTAQITDQEL 1787
QY 91 -----VVOT-----R-----Q-----QA-----98
Db 1788 VELLILAYAPARVHTDWFYDRPESVPADATOGIHDYKQLVRDIAEVNQACAGDLIKS 1847
QY 99 -Y--R-D-K-----L-AQ-----Q-----H-----107
Db 1848 ELIQNRVDCKEYINNHLFKVQKTKLKPGLISGLKVLGISTQSYVDVSHVIGVIRTNQL 1907
QY 108 -----A-A-----A-----AAA---AAA---116
Db 1908 WCTCVTEKCAKATKIKVDLYRKNTGSLRALFGIPRWFDRDITLLEKATAVDAIAGWG 1967
QY 117 -AT-----NO-----OG-----S-----123
Db 1968 KSTEIVKLVNDOCTVVAQTSAAVSNILEKLEQGGKQMKCKVMSIEKMTQQVNTPTLVLD 2027
QY 124 -A-----KN-----G-----E-----128
Db 2028 EASMITWETLSLTGQVENVLYNGNTLQICVLDWYRTGSRATKSILOQAGIIRHYTT 2087
QY 129 --N-----T-A--N-----G--BE-----135
Db 2088 HRIGNPLARELSLVTKELTNAKHETNFTCKSWDAVRWAELTSIAGSLEBPVILCFYNNA 2147
QY 136 --N-----G--A-----H-----T-----140
Db 2148 VRVNMVYKVGCRVDTIHKFQGLEADNVILWCPTGPTGRITLDKHQCLSAATRAKN 2207
QY 141 -I-----ANN-----DM---N-----149
Db 2208 LVWISVNEISNNVPLHKRMGATIGGSKTQPTDENNLADKTLQVSHLSKVMYISKDKT 2267
QY 150 -----E-----VD-----G-----153
Db 2268 IERLECSRLQSQBEGNLVMLKNTLTHLDVYVSGNDHSLRQSGSKLNEKFKMSLSLFS 2327

QY	154	----	D	V	----	E	I	----	157																																								
Db	2328	KEPAPDHVSSSTLGKAEI	V	Q	D	A	N	A	N	K	R	I	N	S	O	I	M	N	I	T	T	Y	Q	K	N	L	A	I	S	G	S	T	A	A	A	T	P	V	L	2387									
QY	158	----	P	----	S	----	N	-	K	----	A	W	----	L	R	----	166																																
Db	2388	TAMLAQPGITKAECF	D	E	P	T	E	T	T	A	N	I	L	M	G	V	N	L	K	I	Q	I	D	F	R	N	D	K	V	I	G	L	S	G	L	R	S	F	V	A	D	2447							
QY	167	----	CH	----	E	S	----	E	-	V	-	F	-	I	C	----	175																																
Db	2448	SSEIKOLYNQKFNCC	G	H	Y	L	K	R	N	F	T	E	I	S	V	K	K	M	N	F	E	R	V	L	M	I	N	V	F	G	V	C	L	S	G	S	L	H	I	S	V	2507							
QY	176	----	AW	----	N	-	P	----	V	----	S	----	182																																				
Db	2508	TIENELYMQSYGGC	S	L	C	G	G	W	I	F	H	N	K	P	I	I	V	G	D	P	S	T	G	N	S	R	T	V	R	F	N	S	K	Q	S	D	K	C	G	P	I	L	A	2567					
QY	183	L-L-V-SGS	----	G	D	S	-	T	A	----	R	I	----	W	----	196																																	
Db	2568	YLRLLNVD	S	G	T	G	V	K	L	P	Y	D	P	P	G	R	G	F	D	M	F	E	G	D	N	V	M	A	A	L	I	E	R	V	L	K	G	P	M	M	A	A	K	L	N	M	L	N	2627
QY	197	----	NL	----	SE	----	200																																										
Db	2628	TNODLCNFDNNSPE	L	F	D	S	L	K	A	L	E	S	G	N	F	T	N	T	L	V	N	E	F	V	G	T	R	L	E	F	N	G	T	N	F	N	T	E	L	Y	N	2687							
QY	201	NS	----	T	-	S	G	----	P	----	T	----	207																																				
Db	2688	VNSNDLRCRPI	C	T	L	D	K	R	Y	L	A	L	M	D	O	N	L	I	W	T	O	I	S	G	K	P	I	H	T	E	A	G	T	R	V	S	N	K	A	F	T	M	A	V	R	2747			
QY	208	Q	----	I	V	----	L	----	R	----	H	C	----	I	-	R	-	E	Q	----	218																												
Db	2748	QMLNRLEFKLT	Q	L	S	R	D	K	P	M	A	N	P	R	G	P	L	L	K	F	M	A	C	H	E	A	D	N	D	K	V	A	A	R	K	Y	K	G	T	O	A	M	2807						
QY	219	----	G	----	Q	----	D	----	221																																								
Db	2808	KLKSRMYLTAWV	N	G	N	S	E	A	S	I	M	Q	H	N	I	T	T	S	N	P	E	N	N	M	L	G	L	V	L	Y	V	A	Q	A	I	H	N	T	G	Y	2867								
QY	222	----	V	----	S	----	N	-	K	----	D	-	V	----	228																																		
Db	2868	SALYITNNACTA	V	L	C	G	H	W	D	F	S	C	P	P	D	G	W	S	T	R	F	L	S	T	H	D	T	I	N	L	A	D	I	K	S	I	L	D	S	E	A	K	L	D	2927				
QY	229	----	T	----	S	----	L	-	D	-	W	----	NS	----	235																																		
Db	2928	QBELSBEETA	V	L	K	E	K	K	E	R	S	E	R	I	H	N	Q	L	R	T	E	D	S	W	Y	S	E	Y	Q	N	V	R	G	T	V	I	V	S	N	C	G	I	S	T	E	S	2987		
QY	236	----	E	-	G	----	T	L	----	T	G	S	----	Y	-	245																																	
Db	2988	IEKIMEITGAE	M	L	C	M	V	P	T	L	P	S	D	N	T	T	A	H	R	V	G	T	G	S	N	N	K	I	T	P	G	A	R	N	I	L	T	P	N	P	E	I	V	C	M	3047			
QY	246	----	D	G	F	----	A	----	R	----	250																																						

Qy	280	--F--I-----LSA--GVD-K-----T-TII-----	292
Db	3407	ITFRARQSIIRRAQTSRLRSYLDNNWEMVKYSALEBAG-DLKVMVHELMRSRIAIKSTGP	3465
Qy	293	-W-----D-----A-H--T--GEA--KO--Q-----F-----	304
Db	3466	TWHLKEDLEAMLDPNLTFAEHLEITKLGDVAVSRWESQKAEKQNEVDFIKSDLGLST	3525
Qy	305	-----P-----	305
Db	3526	LNDHDLNIVLSLLADRPQKVKKDLLTNTLGSQMSVNEAVMLRRKIROQCWESASNCQOL	3585
Qy	306	-----PH-----S--A-----	309
Db	3586	VDHHDHANSVSTTEQOQKLNLSLHBEAGICDINLSDASPLKTRASSYLPEAITTCVQ	3645
Qy	310	-----P-----ALDVD--W-----QS-----	318
Db	3646	TPGMYELTSKPEBIIYEAMDNDECVWRCIEKVYQNIQEPHFRTINGLRAIMOQSKMLTES	3705
Qy	319	-----N--N--T-----F-----AS--C--STD-----	328
Db	3706	QAIIVCOLLGLNCLIQNGETGTVYVFNAPKNPQVQLMRLSODASYDHCVLINLLGADGVK	3765
Qy	329	-----M--C--I-----HV-----C-----	334
Db	3766	RLSPENTAKENMQLEHVCVDNKPVAVGENPYACVSHVIDNEDLARLTSCQPYEKI	3825
Qy	335	-----K--	335
Db	3826	PVLDRPEMLPVANNVWARMTRKTSSATLLKKVTPITGAGSLATRNYYKATPMVTTKPG	3885
Qy	336	-----L-----G--Q-----	338
Db	3886	DVLVLRDHCKKVMCTHTTIVDPGGVEQLLEITDDIGDGAVALDLGLHCLTPADRTPR	3945
Qy	339	-DRP-----IKT--F-----QG-----H-----T 349	
Db	3946	RDRPVNTSNEWLSNIETQRYLEATKRLRSHVGGEGKLIHYHFDNREHHMYDEHDYIDA	4005
Qy	350	-----N--E-----VN-----A--IK-W--D-P-----	359
Db	4006	MLDENIVIELDPTVOVPTGVNIVNLICIRMFRLMLYGIPTWVAYCDDNPSAAQLAWFL	4065
Qy	360	-----	359
Db	4066	KYKMGHKVFNNVTTEDLTNTRAMEIINALFSTYQHTTNTTTRVHLRNKAETILSKEK	4125
Qy	360	-----T--G-----N-----L-----L--L-- 364	
Db	4126	VSEELFNTTIDISVFTAVLMLGENVLKSNQGTCTVDLIPETPEPQLODMLMESGILV	4185
Qy	365	-----A-----S--CS-----DD-M-----	371
Db	4186	DIKQCAVHVPTGTKVALRVVESLSHEIACSTKVGVDPKDDERAGASNLTSQSPASDK	4245
Qy	372	-----T-----L-----KI 375	
Db	4246	ADPGSQLSCSEENMSGTQTPVACVHI EPCDHAANASNDVDLKEITQPEQNNDHIISKI	4305
Qy	376	-----W-----S-----M----- 378	
Db	4306	LATAQAKSNPDKSWKSGATSADLQAIQSLCDSPDGWSQVADQAKKLEVMNFSDEVT	4365
Qy	379	-----K-Q-----D----- 381	
Db	4366	LSGIHQRYNVNSQKYLQVRVDISERWAIYKIGIEHGSVDPTNNPGWMLAFPPMKLGL	4425
Qy	382	-----NC-----V-H-----D----- 386	
Db	4426	VYKEHEIKVTWVYAGNCAYRLNRKIFRKPESVFTHVNVKPEVMDLILTYTSANNKOYPTN	4485
Qy	387	LQ--A-----H-----N-----K-E-I-----Y----- 395	

Db 4486 LQRLADKYLHGRVSHVYVGGYNRDLQTCVELTKRMERVIVVTDAPDVNQLDADIQV 4545
QY 396 ---TK---W---SP---TG---P--- 404
Db 4546 REKIITIKVTSQPEYPYRAIWMRFILPANIDPECELYILNGREMAPNDMELVDRVMEESF 4605
QY 405 ---G---T---N---N--- 408
Db 4606 EDFDIVSALHCOAFNVCCGYAKYNHAKFTDYNTFNDIYEAGNSYGGDEIYMSRYKYKMM 4665
QY 409 ---P-NA---NML---SASF--- 420
Db 4666 WLYYILPGGAWYGTEGKWLNVALDITICLBKEGERTRIDRVKGLFCSANFFHVAPDTKG 4725
QY 421 --D-STV--- 426
Db 4726 IYDSTVDMSVDGIPWQYAMQINGHSVTENPVNRHSTGSGTGNPNRISNDGKPPWQLNPN 4785
QY 427 -W-D---V---D---R-G---I---C--- 434
Db 4786 NWLLDPGLKSKWKIKQFLDSSSVKCIKESQSLPIQLGFLKPIVVPFANQACEATGVSM 4845
QY 435 ---H-TL--- 438
Db 4846 NNTGTLYVEVNIQAQKLAARNEPLQHOTLILKSHLTLMSTTGRHASKWPEYTHLMKADLQ 4905
QY 439 ---K---H-Q---E---P---V--- 445
Db 4906 NCQVYVEGSAITLQWKRETOHAQVYTTBERDVFRIPQALDHSVGTLTATKLVKVTGNE 4965
QY 446 --XS---V---A---F---S---P--- 452
Db 4966 LGLSSSGVLARVLKNDRLAHHWLETOHLQKLEFDRRIISHESSGWTFKPRIGQT 5025
QY 453 ---D---V---H---I---W--- 459
Db 5026 LYLADVLTAAGIECVNHITTSREWKEALKNPLTKVGIMPLTNSVLVSGKLGCVQPIPR 5085
QY 456 -Y---N---L---AS 459
Db 5086 CYAMLCQGVKLIKFKPETRGGTGPSNVANVGSNPPTAGGDPPTPLITSWNKWSGAS 5145
QY 460 ---G-S---FD---K--- 464
Db 5146 NQSPFGVSAAKAYQDKREKKIETPDQLRLVLETMFLDPAVALMSHGLTDLRKLDNGGNWH 5205
QY 465 C---V---H---I---W--- 469
Db 5206 CATDATMKSDFEIIIEPVGGDLLPSLNTSALRHNETNARVIDLWDDTDLRDMWLTLYAPKNP 5265
QY 470 ---N---TQ---TG--- 474
Db 5266 MKLTSRVSPQCKINLKLTLTNRPCTRPVPTQVMGNGENAVTGRLGSLVPLRREPNNVT 5325
QY 475 ---AL--- 476
Db 5326 HELHKFTAYYRGWERVLKDFKANTITISDADVKTWLSRRSDWKALATSTIKMLTGLP 5385
QY 477 ---VH--- 478
Db 5386 SNPMNAVNVHVKTESLLKANPIYWRQTQGRIIWQPKELCALMSPAFIAIKERLKEVLR 5445
QY 479 ---S---Y--- 480
Db 5446 DEIVYTDGLTPDMLSAARARTIQDYVYFEDDLVIQDROTDQELIDLEFQVWMLDGLDNLIA 5505
QY 481 ---R---G---TG--- 484
Db 5506 NLWRLVHNKVRFKQGSWGLDAMRLTGQATTALGNAINTLNVHSSGFVIEHROAIKLMFV 5565
QY 485 -G--I-F---E---V---C--- 490

Db 5566 LGDDNITFMSAEPNLTKYKRLMSERYNMRSKPQVRNVGTFCSLCYRNSFGHCVEGPDF 5625
QY 491 --- 492
Db 5626 VLRHRFEVTGGNHVELPVAVKMPHNVRIGTKFFNGRYLHVHNEBEAEMPEVWELRNH 5685
QY 493 A---A---G---D---K---VCAS---A--- 502
Db 5686 ANRSYILVLKTNSEWRLDGVQVHDLSDKIDTSGNLAEALLEGASGMGNPMPMAAGMSY 5745
QY 503 ---S---D---G---S---VCV---L-D---L---RK 514
Db 5746 LHMGLKSEADKINTEKEFGLERLKWYDVASCKNATCIKYNMSPMELEDNIGELLKMKIRK 5805

RESULT 15
S58870
reelin precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Dec-1996 #sequence revision 06-Dec-1996 #text_change 21-Jul-2003
C;Accession: S58870; S71844; I49297
R;D'Arcangelo, G.; Miao, G.G.; Chen, S.C.; Soares, H.D.; Morgan, J.I.; Curran, T.
Nature 374, 719-723, 1995
A;Title: A protein related to extracellular matrix proteins deleted in the mouse mutant
A;Reference number: I49297; MUID:95231649; PMID:7715726
A;Accession: S58870
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-3461 <DAR>
A;Cross-references: EMBL:U24703; NID:g902486; PID:g902487
R;D'Arcangelo, G.
submitted to the EMBL Data Library, April 1995
A;Reference number: S71844
A;Accession: S71844
A;Molecule type: mRNA
A;Residues: 1-215, 'N', 217-1905, 'S', 1907-3355, 'V', 3357-3391, 'N', 3393-3461 <DA2>
A;Cross-references: EMBL:U24703; NID:g902486; PID:g902487
F;1-27/Domain; signal sequence #status predicted <SIG>
F;28-3461/Product; reelin #status predicted <MAT>
F;1769-1795/Domain; EGF homology <EGF>

Query Match 70.3%; Score 2579.8; DB 2; Length 3461;
Best Local Similarity 11.7%; Pred. No. 1.8e-52;
Matches 390; Conservative 95; Mismatches 16; Indels 2842; Gaps 359;

QY 1 M-----SI-----S-----S-----D-----EV----- 8
Db 1 MERGCWAPRALVLAALLATLRARAATGYPRFPFLCTHHELEGGEGEVLISL 60
QY 9 ---N-----F---LV---VRY---LQ---ES-----GF---S-H 23
Db 61 HIAGNPTVYVPGQYHVHTISTSTFFDGLLVLTG-LYTSISQSSQSIGSSAFGFGIMSDH 119
QY 24 -----S-A-----F-----T-----F-----G-I----- 30
Db 120 QFCNQFMCVWASHVSHLPTNLSFVNIAPPAGTGCWFMATATHRGQVIFKDALAQQLC 179
QY 31 -----E---SH-----I-----S-Q-----SNI-----N-----GAL----- 43
Db 180 EQGAPTEATAYSHLAIHSDSVILRDDFDSDYQQLNPNIVWECNCMEGECQGTIMHGN 239
QY 44 -V-----P-P-----A-AL---I-----SI-----IQ--- 54
Db 240 AVTFCEPYGPRELTTTCLNTTASVQLQFISGSGSCRFSDPSITVSYAKNNTADMIQLE 299
QY 55 -----KG---LQV-----V-BA-E-----VGIN----- 67
Db 300 KIRAPSNVSTVHILYLPBEAKGESVQFQWKQDSLVRGVEYACWALDNLIV-INSARE 358
QY 68 -----ED-----GT-LF-----DG-----R----- 76
Db 359 VVLEDNLDVDPDGNLWFFPGATVKHSCQSDGNSIYFHNESGSEFNATTRDVLSTEDIQ 418

QY 77 -----P-----I-----ES-LSLI-----DA-----86
Db 419 EQWSEFESOPTGWDILGAVVGADCGTVESGLSLVFLKDGKRLCTPYMDTTCYGNLRFY 478
QY 87 -VM-----P-----D-----VV-----QT-----94
Db 479 FVGGICDPCVSHENDIILYAKIEGRKEHIALDTLTYSYKYVPSLVSVVINPELQTPATK 538
QY 95 ---RQ---QAY-R---D-----K-LAQ-----Q-----106
Db 539 FCLROKSHOQYNNRNVMAVDFHVLVPLVPTSMHMIQFSINLGCCTHQPQGNVSLEFSTNH 598
QY 107 ---HA---A---AAA-AA---AA---AA-----AT---118
Db 599 GRWSLLHTECLPEICAGPHLPSTVYSSYNSYGNWRITIPALNALTTRDTRIRWQTGP 658
QY 119 ---N-----Q-----Q-----GS123
Db 659 ILGNMAIDNVVIGPSCFKFCGRGQCTRHGCKDPCGFGPACEMASQTPPMFISESFGS 718
QY 124 AK-----N-----G-E-----N-----129
Db 719 ARLSYHNFYSIRGAEVSGCVLASKALVFNKGRRLQITSLDSSQSRLFTLRIG 778
QY 130 ---T-A---NG---E-E---N-----G-137
Db 779 SKSVLSTCRAPDQPGEGVLLHYSDNGITWKLLEHYVYNYHEPRIISVELPDDARQFGI 838
QY 138 ---A-H---T-----I---AN-----N-----H-145
Db 839 QFRWQPYHSGQEDVWAIDEIVMTSVLFSNISLDFNLVEVTQSLGFLGNVQPYCGHD 898
QY 146 ---T-D---M-----MEVDG-----D-V-E-----156
Db 899 WTLCTGDSKLASSMYETVQSMQI-GASYMIQFSLVMGCGQKYTPHMDNQVLEYSANH 957
QY 157 ---I---PSN---K---A---VVL-----R-----166
Db 958 GLTWHLVQEBCLSPSCQFTSNIYHASEFTQWRVTVVLQKWTSGATRRWQSYY 1017
QY 167 ---GH-----E-SE-----V---P---173
Db 1018 TAQDEWALDNIYIQOCPNMCNSGHGCDHGVCRCDOGYQGTCHPEAALPTSIMSDPENP 1077
QY 174 ---I---C-----A-----W-----177
Db 1078 SSWEDMQEVIGGEVWKPEQCGVYSSGSSLYFSKAGKQLVSWDLDTSWVDVQFYQI 1137
QY 178 ---N-P---V-----SDL---LV-----SG-----187
Db 1138 GBESAACNKDPSREEGILLQYNSNGGIQWHLLEMYFSDFSKPRFVYLELPAAGKTPCTR 1197
QY 188 ---SG---D---S-----T-----A-----R-IW-196
Db 1198 FRWKPVFSGEDYDQWAVDDIILSEKQKQVIVPVNPTLPQNFYKPAFDYPMNQMSVWL 1257
QY 197 ---N-ISEN-S---T-----S-G-----P-----T-QL---209
Db 1258 MLANEGMAKNDSCATTPSAMVFGKSDGRFAVTRDTLKPQVYVLPKLNIGCTQSFSST 1317
QY 210 ---V-LR-H---C-----IR-EG-----QOD-----V222
Db 1318 APVLLQYSHDAGMSWFLLEKCCFPASAAKCEGNSRELSEPTVYTYG-DPEEWTRITIAI 1376
QY 223 P-----S-N-----K-----D-----227
Db 1377 PRSLASKTRFRWIOESSQXNPPFGLDGYVISEPCPYCSGHGDCISGVCFDLGYTA 1436
QY 228 ---V-T---S-ID-----W-----NSEG-TL---L---A-241
Db 1437 AQGTCVSNTPNHSEMFDRFEGKLSPLMYKITGGQVGTGCGTFLN-DGRSLYFNLGKREAR 1495
QY 242 ---T-----GS-----Y-D-G-----F248

Db 1496 TVPLDRNISLVQFYIQIGSKTSGITYITPRARYEGLVVQYSGNDNGILWHLRLDFMSF 1555
QY 249 -----A-RW-----T---K-DG-256
Db 1556 LEPQIISIDLPREAKTPATAFR-WMOPQHKHSAQWALGDVLIGNVDSQTFQDKDDGS 1614
QY 257 ---N-----LAS-T-L---G-Q---H-----KG-267
Db 1615 IDLOANWYRIQGOVDIDCL-SMDTALIFTENIGNPRYAETWDFHVSESFLOWENMGC 1673
QY 268 ---PIF-A---L-----K-W---NKK-----G-----278
Db 1674 SKP-FSAGHIGIQLOYSNLNGKDWLVBECVPTTICGVHYTESSTYTSERFQNRRTVY 1732
QY 279 ---N-----P-I---LSAGVDKT-TI---I---W-----DA-----295
Db 1733 LPLATNSPRTFRWITQNTYTVGAD-SWADNVILASGCPMCMCGRIGICDSGRVCVDRGFG 1791
QY 296 ---H-----H-----T---GE-----299
Db 1792 GPFCVPVPLPSILKDDFNGNLHPDLWPEVYGAERGNLGETIKSGTCLIFKGEGLRMLI 1851
QY 300 ---AK-Q---QF-----P-----F-----H-----SA-309
Db 1852 SRDLCTNTMYQVQSLRFIAKGTPEPERSHISILLOFVSQVGVTHLMDDEFYPPQTRILFIN 1911
QY 310 ---P---A---L---D-----V-----D-W316
Db 1912 VLPYGAQTNATFRFLWQPYNNGKKEEIIIDDFIIDGNLNNPVLILLDTDFGPREDNW 1971
QY 317 -----Q-S-N-NT-----321
Db 1972 FFYPGNGIGLYCPSYKSGAPEDSAMVFSVNEVGEHSITRDLNVNENTIIQFENVGCS 2031
QY 322 ---F-A-----S-CST-----327
Db 2032 TDSSADPVRLEPFRDQFATWHLPLCYHSSSLVSSLCSTHSPSTYVAGTTOGWRE 2091
QY 328 -----D-----MC-IH-----VC334
Db 2092 VHFGLHLGCVSRFRWYQGFYAGPQVTVWALDNVYIQQCEMYGHCSCINGTKCIC 2151
QY 335 ---K---L---GQ---DR-----P-----I-KT-----P-Q-G-347
Db 2152 DPCYSGFTCKISTKNPDFLKDDFEGQLESDFLLMSGKPSKCGILSSGNLFFNEDGL 2211
QY 348 ---H-----T-N-----EV-----352
Db 2212 RMLVTRDLDSLHARFVQFFMRLCGKGVDPDRSQPVLLQVSLNGLSWSLLOBFLEFSNS 2271
QY 353 NA---I---K-----W-----357
Db 2272 NVGRYIALEMPKARSSTLRWQPSENGHFYSPWVIDOILIGGNISGNTVLEDDPSTL 2331
QY 358 D---P---TG-----N---IL-----ASCS--D369
Db 2332 DSRKWLHPGCTKMPVCGSTGDALVIEKASTRYVVTDDIADVNEDSLQIDFAASCSTVD 2391
QY 370 ---DM-----T---L-KI-----W-----376
Db 2392 SCVAIELEYSDGLSWHPLVRDCLPTNVESCRYHLQRIILVSDTFNKRWTRITLPLPSYTR 2451
QY 377 S-----M-----KQ---DN---CV-----H-----D---386
Db 2452 SQATFRWQAPDFDKQWALDNVYIGDCCLDMCGHRCVQSGCVCDQWGLYCDPE 2511
QY 387 ---L---Q-----A-H-----390
Db 2512 ETSLPQLKDNFNRAPSNQNLVTVSGKLSVCGAVASGLAHFSGCSRLVTVDLNLT 2571
QY 391 NKE-I---Y---T-----IKW-----S-PTG-----P404

Db 2572 NAEFIQFYPMYGLITPSNRNOGVLLSEYVNGGITWNLLMEIFYDOYSKP-GFVNILLPP 2630
QY 405 -----CT-----NNP-----N--A--NLMLA-SA-----SF-----D 421
Db 2631 DAKIATRFRRWQ-PRHDLQDNDWADINVLISGSADQRTVMLDTFSSAPVQHERSPAD 2689
QY 422 S-TV-R-----L-----W-----DVD-----R-----GI--C-IH-----TL--TKH- 441
Db 2690 AGFVGRIAFEMFLEDKTSVNEENLPHD-DCTVERFCDSPDGVMLCGSHDGREVYAVT-HD 2747
QY 442 -----Q-----E-P-----V-YSAF-----S-----P-D----- 453
Db 2748 LPTENIMQFKISVCKPEKIAQNIHVQFSTDFGVSWYLVPOCLPADPKCGSVSQ 2807
QY 454 -----G-R-Y-L-AS-G-----S-----F-----D- 463
Db 2808 PSVFPTEGWKRITYPLPSLTGNPVRFRFYQKYSVQWAI DNFLPGCLDNCGGHDC 2867
QY 464 -K-CV-----H-----I-----W-----NT- 471
Db 2868 LKEQCICDPGSGPNCYLTHSLKTLKRFDFSEIKPDLWMSLEGGSTCTCGVLAENTA 2927
QY 472 -----Q-T-----GA-L-----V-----H-----S----- 479
Db 2928 LYFGSTVRQAITQDLRLGAKFLQYWGRIIGSENNTSCHRFPVCRKEGVLLDFSTDGKIT 2987
QY 480 -----Y--R--G-----T-----G-----GI----- 486
Db 2988 WTLLEMDFKYISVRHDYILLPEGALTNTTLRWQPFVISNGLVVGVCACAVGTQGH 3047
QY 487 -----F-----E-----V-----C-----W-----NA 493
Db 3048 SDWWSRNHPSQLVDTFDEGSSHEENWSFYPNVARTAGFCGNPSFHLWPNKKKDKTHNA 3107
QY 494 -----A-----GD-----K----- 497
Db 3108 LSSRELIQPGYMMQFKIVVGCEATSCGDLHVMLEYTKDARSQSWQLVQTOCLPSSNS 3167
QY 498 VG-----A-----SA-----S-----D----- 504
Db 3168 IGCSPQFHEATLYNAVSSWKRIITQLPDHVSSATOFRWIQKGEETEKOSWAI DHVY 3227
QY 505 -G-----S-----VC----- 508
Db 3228 IGEACPCLCSGHGYCTTGAVCICDESFGQDDCSVFSHELPSYIKDNFESARVTEANWETI 3287
QY 509 ----V-----LDL-R--K 514
Db 3288 QGGVIGGCCQLAPYAHGDSLYFNGCQIRQAATKPLDLTRASK 3330

Search completed: January 3, 2005, 15:29:45
Job time : 114.667 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:55:11 ; Search time 102.333 Seconds
(without alignments)
2889.993 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 3669

Sequence: 1 MSISSDEVNFLVRYLQESG.....GDKVGASADGSCVCLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3665	99.9	514	1 TBLR MOUSE	Q8bj5 mus musculus
2	3622	98.7	514	1 TBLR HUMAN	Q9bzk7 homo sapien
3	3583.5	97.7	519	2 Q7S2M9	Q7szm9 xenopus lae
4	3577	97.5	522	2 Q6GPC6	Q6gpc6 xenopus lae
5	3397.9	92.6	527	1 TBLX MOUSE	Q9qxe7 mus musculus
6	3350.2	91.3	526	1 TBLX HUMAN	Q60907 homo sapien
7	3247.4	88.5	522	1 TBLX HUMAN	Q9bq87 homo sapien
8	2937.5	80.1	511	2 Q7Q371	Q7q371 anopheles g
9	2916.9	79.5	700	1 EBI DROME	Q95tj9 drosophila
10	2827.8	71.6	5636	2 Q96RW7	Q96rw7 homo sapien
11	2822.8	71.5	20925	2 Q8NUX1	Q8njx1 trichoderma
12	2620.7	71.4	7073	2 AAP49011	Aap49011 sars coro
13	2618.9	71.4	7191	2 Q6AXA9	Q6xa09 alternaria
14	2618.9	71.4	7191	2 AAP78735	Aap78735 alternari
15	2612.9	71.2	4599	1 LR1B MOUSE	Q9ji18 mus musculus
16	2612.1	71.2	5175	2 Q8IOU3	Q8io13 caenorhabdi
17	2611.2	71.2	34350	2 Q8W442	Q8w442 homo sapien
18	2610.8	71.2	5198	2 Q7P518	Q7p518 caenorhabdi
19	2610.4	71.1	7105	2 Q7PXX9	Q7pxw9 anopheles g
20	2610.1	71.1	4796	2 Q9NL88	Q9nl88 drosophila
21	2609.7	71.1	7073	1 R1AB CVHSA	P59641 h replicase
22	2609.7	71.1	7073	2 Q6JH48	Q6jh48 sars corona
23	2609.7	71.1	7073	2 Q6R7Y8	Q6r7y8 sars corona
24	2609.7	71.1	7073	2 Q6RCW7	Q6rcw7 sars corona
25	2609.7	71.1	7073	2 Q6RCX8	Q6rcx8 sars corona
26	2609.7	71.1	7073	2 Q6RCY9	Q6rcy9 sars corona
27	2609.7	71.1	7073	2 Q6RD00	Q6rd00 sars corona
28	2609.7	71.1	7073	2 Q6RD11	Q6rd11 sars corona
29	2609.7	71.1	7073	2 Q6RD22	Q6rd22 sars corona
30	2609.7	71.1	7073	2 Q6RD33	Q6rd33 sars corona
31	2609.7	71.1	7073	2 Q6RD44	Q6rd44 sars corona

Q6rd55 sars corona
Q6rd66 sars corona
Q6rpe9 sars corona
Q6uzf1 sars corona
Q6uzf5 sars corona
Q6v586 sars corona
Q6va80 sars corona
Q6va91 sars corona
Q6vaa2 sars corona
Aar14802 sars coro
Aar14806 sars coro
Aar14810 sars coro
Aap82978 sars coro
Aar91584 sars coro

ALIGNMENTS

RESULT 1
TBLR MOUSE STANDARD; PRT; 514 AA.
ID Q8BHJ5; Q8CEG4; Q8VEG3; Q9EQD4;
AC Q8BHJ5; Q8CEG4; Q8VEG3; Q9EQD4;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE P-box-like/WD-repeat protein TBLR1 (Nuclear corepressor/HDAC3
DE complex subunit TBLR1) (TBL1-related protein 1).
GN Name=Tblr1; Synonyms=Iral;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Zhang X., Dromady S., Basch R.;
RT "Identification of four human CDNAs that are differentially expressed
RT by early hematopoietic progenitors."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Testis;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Baissel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Mulcais L., Marchionni L., Mckenzie L., Miki H.,
RA Nagahina T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sautelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski B., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).

[3]
 RP SEQUENCE OF 314-514 FROM N.A.
 RC TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Bouffard G.G.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RL and mouse cDNA sequences.";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: F-box-like protein involved in the recruitment of the
 ubiquitin/19S proteasome complex to nuclear receptor-regulated
 transcription units. Plays an essential role in transcription
 activation mediated by nuclear receptors. Probably acts as
 integral component of the N-Cor corepressor complex that mediates
 the recruitment of the 19S proteasome complex, leading to the
 subsequent proteosomal degradation of N-Cor complex, thereby
 allowing cofactor exchange, and transcription activation (By
 similarity).
 CC -1- SUBUNIT: Component of the N-Cor repressor complex, at least
 composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.
 CC Probable component of some E3 ubiquitin ligase complex. Interacts
 CC with histones H2B and H4 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC apparently displays the same function as component of ubiquitin E3
 CC ligase complexes (By similarity).
 CC -1- SIMILARITY: Belongs to the WD-repeat EBI family.
 CC -1- SIMILARITY: Contains 1 F-box-like domain.
 CC -1- SIMILARITY: Contains 1 Lish domain.
 CC -1- SIMILARITY: Contains 8 WD repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL; AF268195; AAG44738.1; -
 CC ENBL; AK029595; BAC26526.1; -
 CC ENBL; AK033347; BAC28241.1; -
 CC ENBL; AK036064; BAC29294.1; -
 CC ENBL; BC018512; AAH18512.1; -
 CC HSP; P16649; 1ERJ.
 CC InterPro: IPR006594; Lish.
 CC InterPro: IPR011047; Quin_alc_DH_like.
 CC InterPro: IPR001680; WD40..
 CC Pfam: PF00400; WD40; 8.
 CC PRINTS: PR00320; GPROTEINRPT.
 CC ProDom: PD000018; WD40; 3.
 CC SMART: SM00667; Lish; 1.
 CC SMART: SM00320; WD40; 8.
 CC PROSITE; PS00896; LISH; 1.
 CC PROSITE; PS00678; WD_REPEATS_1; 4.
 CC PROSITE; PS00082; WD_REPEATS_2; 6.
 CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
 CC Activator; Nuclear protein; Repeat; Transcription regulation;
 KW

KW Ub1 conjugation pathway; WD repeat.
 FT DOMAIN 4 36 Lish.
 FT REPEAT 41 86 F-box-like.
 FT REPEAT 167 206 WD 1.
 FT REPEAT 223 262 WD 2.
 FT REPEAT 264 303 WD 3.
 FT REPEAT 306 344 WD 4.
 FT REPEAT 347 386 WD 5.
 FT REPEAT 389 437 WD 6.
 FT REPEAT 440 479 WD 7.
 FT REPEAT 481 513 WD 8.
 FT DOMAIN 108 117 Poly-Ala.
 FT CONFLICT 185 185 A -> V (in Ref. 1).
 FT CONFLICT 374 374 K -> R (in Ref. 2; BAC29294).
 FT CONFLICT 480 480 Y -> C (in Ref. 3).
 SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;
 Query Match 99.9%; Score 3665; DB 1; Length 514;
 Best Local Similarity 99.8%; Pred. No. 1.4e-65;
 Matches 513; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISIIKGLQVY 60
 DB 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISIIKGLQVY 60
 QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQQHAAAAAATNQ 120
 DB 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQQHAAAAAATNQ 120
 QY 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFIKAWNPV 180
 DB 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFIKAWNPV 180
 QY 181 SLLIVSGSGDSTARINLSENSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 DB 181 SLLIVSGSGDSTARINLSENSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
 QY 241 ATGSYDGFARIWTKGNLSTLQHGKPIFALKNNKGNFILLSAGVDKTTIWDATGEA 300
 DB 241 ATGSYDGFARIWTKGNLSTLQHGKPIFALKNNKGNFILLSAGVDKTTIWDATGEA 300
 QY 301 KQGFPHSPALPDVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 DB 301 KQGFPHSPALPDVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKIYTIKWSPTGPTGNPNANMLASASF 420
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKIYTIKWSPTGPTGNPNANMLASASF 420
 QY 421 DSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSPDKCVHIWNTQTGALVHSY 480
 DB 421 DSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSPDKCVHIWNTQTGALVHSY 480
 QY 481 RTGGGIFEVCAWNAAGDKVGSASDGSVCVLDLRK 514
 DB 481 RTGGGIFEVCAWNAAGDKVGSASDGSVCVLDLRK 514
 RESULT 2
 ID TBLR HUMAN STANDARD; PRT; 514 AA.
 AC Q9BZK7; Q9H211; Q9H9A1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE F-box-like/WD-repeat protein TBLR1 (Nuclear receptor corepressor/HDAC3
 DE complex subunit TBLR1) (TBL1-related protein 1).
 GN Name=TBLR1; Synonyms=IRAL;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]

SEQUENCE FROM N.A., IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR COMPLEX WITH NCOR1, NCOR2, GPS2, TBL1X AND HDAC3.
MEDLINE=21930350; PubMed=11931768;
Zhang J., Kalkum M., Chait B.T., Reeder R.G.;
"The N-COR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";
Mol. Cell 9:611-623 (2002).
[2]
SEQUENCE FROM N.A.
PubMed=11063877;
Zhang X., Dormady S.P., Basch R.S.;
"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors.";
Exp. Hematol. 28:1286-1296 (2000).
[3]
SEQUENCE FROM N.A.
PubMed=14702039; DOI=10.1038/ngl1285;
Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Obayashi M., Nishi T., Shibahata T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K., Arita M., Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S., Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata H., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Yamazaki M., Watanabe M., Tashiro H., Tanigami S., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M., Togashi T., Oyama K., Hata H., Watanabe M., Komatsu T., Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K., Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R., Nakai K., Yada T., Nakamura Y., Ohara O., Isoqai T., Sugano S.;
"Complete sequencing and characterization of 21,243 full-length human cDNAs.";
Nat. Genet. 36:40-45 (2004).
[4]
SEQUENCE OF 1-125 FROM N.A.
RC TISSUE=Adrenal cortex;
MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Scapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E., Schnur A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[5]

RP COMPONENT OF THE N-COR COMPLEX WITH TBL1X, CORO2A AND HDAC3, AND
RX HISTONE-BINDING.
PubMed=12628926;
Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,
Wong J.;
"Purification and functional characterization of the human N-COR complex: the roles of HDAC3, TBL1 and TBL1L1.";
EMBO J. 22:1336-1346 (2003).
[6]
FUNCTION AND RECRUITMENT OF 19S PROTEASOME COMPLEX.
RP PubMed=14980219;
Riess V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;
"A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors.";
Cell 116:511-526 (2004).
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the ubiquitin/19S proteasome complex to nuclear receptor-regulated transcription units. Plays an essential role in transcription activation mediated by nuclear receptors. Probably acts as integral component of the N-Cor corepressor complex that mediates the recruitment of the 19S proteasome complex, leading to the subsequent proteasomal degradation of N-Cor complex, thereby allowing cofactor exchange, and transcription activation.
CC -!- SUBUNIT: Component of the N-Cor repressor complex, at least composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.
CC Probable component of some E3 ubiquitin ligase complex. Interacts with histones H2B and H4.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and apparently displays the same function as component of ubiquitin E3 ligase complexes (By similarity).
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -!- SIMILARITY: Contains 1 F-box-like domain.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to a potential poly-A sequence starting in position 125.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF314544; AAK00301.1; -;
DR EMBL; AF268193; AAG44736.1; -;
DR EMBL; AK022956; BAB1431.1; -;
DR EMBL; BC060320; AAH60320.1; ALT_TERM.
DR HSSP; P16649; IERJ.
DR MIM; 608628; -;
DR InterPro; IPR01047; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR PRODOM; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Activator; Nuclear protein; Repeat; Transcription regulation;
KW Ub1 conjugation pathway; WD repeat.
FT DOMAIN 4 36 Lish.
FT DOMAIN 41 86 F-box-like.
FT REPEAT 167 206 WD 1.
FT REPEAT 223 262 WD 2.
FT REPEAT 264 303 WD 3.


```

RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Dev. Dyn. 225:384-391(2002).
RN [2]
RN PROC. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RN SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 8 WD repeats.
DR EMBL; BC073215; AAH73215.1; -.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR011047; Quin_alc_DH_like.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 522 AA; 56284 MW; 4FB0F58F503B461F CRC64;

Query Match 97.5%; Score 3577; DB 2; Length 522;
Best Local Similarity 96.0%; Pred. No. 1.le-63;
Matches 502; Conservative 4; Mismatches 7; Indels 10; Gaps 4;

QY 1 MSISSEVNFVLYRYLQESGFSATFTGIESHSQSNGALVPPAALISIIQGLQYV 60
DB 1 MSISSEVNFVLYRYLQESGFSATFTGIESHSQSNGALVPPAALISIIQGLQYV 60
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVPDPVDTQQAQYRDKLAQOH-----AAAAAAA 115
DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVPDPVDTQQAQYRDKLAQOH-----AAAAAAA 120
QY 116 AAT---NQGSAGKNGENTANGEAGHTIANNHTDMWVDGVEI-PSNKAVVLRGHESE 171
DB 116 AAT---NQGSAGKNGENTANGEAGHTIANNHTDMWVDGVEI-PSNKAVVLRGHESE 179
QY 172 VFTCAWNPVSDLLVSGSGDSTARIWNLSNSTSGTQQLVLRHCIRGGQDVPNSKDVTSL 231
DB 172 VFTCAWNPVSDLLVSGSGDSTARIWNLSNSTSGTQQLVLRHCIRGGQDVPNSKDVTSL 239
QY 180 VFTCAWNPVSDLLVSGSGDSTARIWNLSNSTSGTQQLVLRHCIRGGQDVPNSKDVTSL 239

```

RESULT 5

TBLX_MOUSE

ID TBLX_MOUSE STANDARD; PRT; 527 AA.

AC Q9QXE7; Q8BMM0; Q8BYQ4; Q8COA1;

DT 16-OCT-2001 (Rel. 40, Created)

DT 05-JUL-2004 (Rel. 44, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE F-box-like/WD-repeat protein TBLX (Transducin beta-like 1X protein).

GN Name=Tblx; Synonyms=Tbl;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RC TISSUE=Hypothalamus, Medulla oblongata, and pituitary;

RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;

RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Hono H., Kondo S.,

RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,

RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,

RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,

RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,

RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,

RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,

RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,

RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,

RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,

RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,

RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,

RA Nagahima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,

RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,

RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,

RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,

RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,

RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,

RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,

RA Yuan Z., Zavalon N., Zhu Y., Zimmer A., Carninci P., Hayashizaki Y.,

RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,

RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,

RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,

RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,

RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,

RA Birney E., Hayashizaki Y.;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F., Diatchenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; et al.
 "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences".
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [3]
 SEQUENCE OF 200-527 FROM N.A.
 STRAIN-129/SwJ;
 Botcherby M.R.M., Straw R., Clarke D., Greystrom J.S., Weston P., Hunter G., Kimberly C., Rhodes M.; et al.
 Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 TISSUE SPECIFICITY.
 MEDLINE=99264241; PubMed=10330347;
 Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A., Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.; et al.
 "X-linked late-onset sensorineural deafness caused by a deletion involving OAI and a novel gene containing WD-40 repeats".
 Am. J. Hum. Genet. 64:1604-1616(1999).
 CC -1- FUNCTION: F-box-like protein involved in the recruitment of the ubiquitin/19S proteasome complex to nuclear receptor-regulated transcription units. Plays an essential role in transcription activation mediated by nuclear receptors. Probably acts as integral component of corepressor complexes that mediates the recruitment of the 19S proteasome complex, leading to the subsequent proteasomal degradation of transcription repressor complexes, thereby allowing cofactor exchange (By similarity).
 CC -1- SUBUNIT: Component of the N-CoR repressor complex, at least composed of NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2. Component of a E3 ubiquitin ligase complex containing UBE2D1, STAH1, CACVBP/SIP, SKP1A, APC and TBL1X. Probably part of other corepressor complexes, that do not contain NCOR1 and NCOR2. Interacts with histones H2B, H3a and H4 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- TISSUE SPECIFICITY: Expressed in the cochlea.
 CC -1- DOMAIN: The F-box-like domain is related to the F-box domain, and apparently displays the same function as component of ubiquitin E3 ligase complexes (By similarity).
 CC -1- SIMILARITY: Belongs to the WD-repeat EBI family.
 CC -1- SIMILARITY: Contains 1 F-box-like domain.
 CC -1- SIMILARITY: Contains 1 Lish domain.
 CC -1- SIMILARITY: Contains 8 WD repeats.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to erroneous gene model prediction.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; AK030547; BAC27015.1; -
 EMBL; AK031937; BAC27612.1; -
 EMBL; AK038674; BAC30092.1; -
 EMBL; BC043105; AB43105.1; -
 EMBL; F38006; CAB61534.1; AUT_SEQ.
 MGD; MGII:1336172; Tbl1x.

DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 5.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR PRODOM; PD000018; WD40; 2.
 DR SMART; SM00320; WD40; 5.
 DR PROSITE; PS00896; LISH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 4.
 DR PROSITE; PS00082; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Activator; Nuclear protein; Repeat; Transcription regulation;
 KW Ub1 conjugation pathway; WD repeat.
 FT DOMAIN 41 36 Lish.
 FT DOMAIN 41 86 F-box-like.
 FT REPEAT 180 219 WD 1.
 FT REPEAT 236 275 WD 2.
 FT REPEAT 277 316 WD 3.
 FT REPEAT 319 359 WD 4.
 FT REPEAT 360 399 WD 5.
 FT REPEAT 402 450 WD 6.
 FT REPEAT 453 492 WD 7.
 FT REPEAT 494 526 WD 8.
 FT DOMAIN 108 121 Poly-Ala.
 FT DOMAIN 130 133 Poly-Ala.
 FT CONFLICT 104 104 A -> T (in Ref. 1; BAC27015).
 FT CONFLICT 349 349 L -> H (in Ref. 1; BAC27612).
 SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;
 Query Match 92.8%; Score 3397.9; DB 1; Length 527;
 Best Local Similarity 87.6%; Pred. No. 6.9e-60;
 Matches 465; Conservative 28; Mismatches 17; Indels 21; Gaps 9;
 QY 1 MSTSSDEVNFLVYRYLQESGFHSATFGIESHSOSNINGALVPPAALISIIKGLQYV 60
 DB 1 MSTSSDEVNFLVYRYLQESGFHSATFGIESHSOSNINGALVPPAALISIIKGLQYV 60
 QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMDVQVTRQOAYRDKLAQOH--AAAAAAAAT 118
 DB 61 EAEVSNEDGTLFDGRPIESLSLIDAVMDVQVTRQOAYRDKLAQOQAAAAAAAAT 120
 QY 119 -----NQGSAKNGENTANGEAHTIANNHTDM-MEVDGDVEI-PSNKAVV 164
 DB 121 ATSTAATTPAAAAQQNPFPKNGEATVNGEENGAI--NNHS-KPMEIDGVEIPPS-KATV 177
 QY 165 LRGHSEVEFCANVPVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGDVPS 224
 DB 178 LRGHSEVEFCANVPVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGDVPS 237
 QY 225 NKDVTSLDWNBSGTLTATSGYDGFARITWKDGNLASTLGQHGKPIFALKWNKKNFILSA 284
 DB 238 NKDVTSLDWNBSGTLTATSGYDGFARITWKDGNLASTLGQHGKPIFALKWNKKNFILSA 297
 QY 285 GVDKTTIIWDAHTGEAKQPPHSPALDWDQSN--TFASCTDNCIHVKLGQDRPIK 343
 DB 298 GVDKTTIIWDAHTGEAKQPPHSPALDWDQ--NNTTFASCTDNCIHVKLGQDRPVK 356
 QY 344 TFOGHTNEVNAIKWDPDTGNLLASCSDMTLKTWSMKODNCVHDLOAHNKEIYTKWSPG 403
 DB 357 TFOGHTNEVNAIKWDPDGMLLASCSDMTLKTWSMKQDACVHDLOAHNKEIYTKWSPG 416
 QY 404 PGTNNPNANMLASAFDSTVRLWDVDRGICHTLTKEPVSVAFSPDGRYLAGSGFD 463
 DB 417 PATSNPNSNIMLASAFDSTVRLWDVDRGICHTLTKEPVSVAFSPDGRYLAGSGFD 476
 QY 464 KCVHTWNTGTGALVHSYRGTTGFEVCWNAAGDKVGASADSGSVCLDRK 514
 DB 477 KCVHTWNTGTGALVHSYRGTTGFEVCWNAAGDKVGASADSGSVCLDRK 527
 RESULT 6
 TELX HUMAN
 ID TELX HUMAN STANDARD; PRT; 526 AA.
 AC O60907; O86UY2;
 DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 05-JUL-2004 (Rel. 44, Last annotation update)
 DE F-box-like/WD-repeat protein TBL1X (Transducin beta-like 1X protein)
 DE (Transducin-beta-like 1, X-linked) (SNAP55).
 GN Name-TBL1X; Synonyms=TBL1;
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DISEASE.
 RX MEDLINE=99264241; PubMed=10330347;
 RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,
 RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.;
 RT "X-linked late-onset sensorineural deafness caused by a deletion
 RT involving OAI and a novel gene containing WD-40 repeats";
 RL Am. J. Hum. Genet. 64:1604-1616(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Lymph, and Pancreas;
 RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Hopkins R.F., Zerbberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Whiting J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Fahey J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR
 RP COMPLEX WITH NCOR2 AND HDAC3.
 RX PubMed=10809664;
 RA Guenther M.G., Lane W.S., Fischle W., Verdin E., Lazar M.A.,
 RA Shiekhattar R.;
 RT "A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-
 RT repeat protein linked to deafness";
 RL Genes Dev. 14:1048-1057(2000).
 RN [4]
 RP COMPONENT OF THE N-COR COMPLEX WITH NCOR2 AND HDAC3.
 RX PubMed=10944117;
 RA Li J., Wang J., Wang J., Nawaz Z., Liu J.M., Qin J., Wong J.;
 RT "Both corepressor proteins SMRT and N-CoR exist in large protein
 RT complexes containing HDAC3";
 RL EMBO J. 19:4342-4350(2000).
 RN [5]
 RP SUBUNIT OF A COMPLEX WITH UBE2D1, CACYBP, SIAH1 AND APC.
 RX PubMed=11389839;
 RA Matsuzawa S.-I., Reed J.C.;
 RT "Siah-1, SIP, and Ebi collaborate in a novel pathway for beta-catenin
 RT degradation linked to p53 responses";
 RL Mol. Cell 7:915-926(2001).
 RN [6]
 RP COMPONENT OF THE N-COR COMPLEX WITH NCOR1, NCOR2, GPS2, TBL1R AND
 RP HDAC3.
 RX MEDLINE=21930350; PubMed=119311768;
 RA Zhang J., Kalkum M., Chait B.T., Roeder R.G.;
 RT "The N-COR-HDAC3 nuclear receptor corepressor complex inhibits the JNK
 RT pathway through the integral subunit GPS2";
 RL Mol. Cell 9:611-623(2002).
 RN [7]

RP COMPONENT OF THE N-COR COMPLEX WITH TBL1R, CORO2A AND HDAC3, AND
 RP HISTONE-BINDING.
 RX PubMed=12628926;
 RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,
 RA Wong J.;
 RT "Purification and functional characterization of the human N-CoR
 RT complex: the roles of HDAC3, TBL1 and TBL1R";
 RL EMBO J. 22:1336-1346(2003).
 RN [8]
 RP FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.
 RX PubMed=14980219;
 RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;
 RT "A corepressor/coactivator exchange complex required for
 RT transcriptional activation by nuclear receptors and other regulated
 RT transcription factors";
 RL Cell 116:511-526(2004).
 CC -!- FUNCTION: F-box-like protein involved in the recruitment of the
 CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
 CC transcription units. Plays an essential role in transcription
 CC activation mediated by nuclear receptors. Probably acts as
 CC integral component of corepressor complexes that mediates the
 CC recruitment of the 19S proteasome complex, leading to the
 CC subsequent proteosomal degradation of transcription repressor
 CC complexes, thereby allowing cofactor exchange.
 CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least
 CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.
 CC Component of a E3 ubiquitin ligase complex containing UBE2D1,
 CC SIAH1, CACYBP/SIP, SKP1A, APC and TBL1X. Probably part of other
 CC corepressor complexes, that do not contain NCOR1 and NCOR2.
 CC Interacts with histones H2B, H3a and H4.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and
 CC apparently displays the same function as component of ubiquitin E3
 CC ligase complexes (By similarity).
 CC -!- DISEASE: Defects in TBL1X may be a cause of late-onset
 CC sensorineural deafness (OASD) [MIM:300650]. OASD is an X-linked
 CC recessive disorder characterized by ocular albinism and
 CC progressive sensorineural hearing loss in the fourth and fifth
 CC decades of life. OASD may be caused by deletion of both GPR143/OAI
 CC and TBL1X adjacent genes; TBL1X defects possibly causing the
 CC hearing phenotype.
 CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
 CC -!- SIMILARITY: Contains 1 F-box-like domain.
 CC -!- SIMILARITY: Contains 1 Lish domain.
 CC -!- SIMILARITY: Contains 8 WD repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y12781; CAA73319.1; ALT_INIT.
 DR EMBL; BC032708; AAH32708.1; --_INIT.
 DR EMBL; BC052304; AAH52304.1; ALT_INIT.
 DR HSSP; P16649; IERJ.
 DR GENE; HGNC:11585; TBL1X.
 DR MIM; 300196; --.
 DR MIM; 300650; --.
 DR GO; GO:0007605; P:perception of sound; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 DR GO; GO:0007601; P:visual perception; TAS.
 DR InterPro; IPR006594; Lish.
 DR InterPro; IPR011047; Quin_alc_DH_like.
 DR InterPro; IPR001680; WD40-
 DR Pfam; PF00400; WD40; 8.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00667; Lish; 1.
 DR SMART; SM00320; WD40; 8.

```
DR PROSITE; PS50896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Activator; Deafness; Nuclear protein; Repeat;
KW Transcription regulation; Ubl conjugation pathway; WD repeat.
FT DOMAIN 4 36 Lish.
FT DOMAIN 41 86 F-box-like.
FT REPEAT 179 218 WD 1.
FT REPEAT 235 274 WD 2.
FT REPEAT 276 315 WD 3.
FT REPEAT 318 358 WD 4.
FT REPEAT 359 398 WD 5.
FT REPEAT 401 449 WD 6.
FT REPEAT 452 491 WD 7.
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 Poly-Ala.
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6E9 CRC64;

Query Match          91.3%; Score 3350.2; DB 1; Length 526;
Best Local Similarity 86.2%; Pred. No. 7.1e-59;
Matches 456; Conservative 36; Mismatches 19; Indels 18; Gaps 10;

QY 1 MSISDEVNPLVRYLQESGFSHTFTGIESHSQSNINGALVPPAALISIIKGLQV 60
DB 1 MSISDEVNPLVRYLQESGFSHTFTGIESHSQSNINGTLVPPAALISIIKGLQV 60
QY 61 EAEVINEGDTLPGDRPIESLSLIDAMPDVVOTROQAVRDKLAQH--AAAAAAA--- 115
DB 61 EAEVINEGDTVDPGRPIESLSLIDAMPDVVOTROQAFREKLAQQAASAAAAATAA 120
QY 116 --RATWQ-QG-S---AKGENTANGENGAAHTIANNHTDM-MEVDGDVEIPSKAVLR 166
DB 121 ATAATTSAGVSHQNPFSKNEATVNGEENRAHSV--NNHA-KPWEIDGEVEIPSKAVLR 178
QY 167 GHESEVFIICAWNPSVOLLVSGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQVPSNK 226
DB 179 GHESEVFIICAWNPSVOLLVSGSGDSTARIWNLSNENSGSTQLVLRHCIREGGHVP 238
QY 227 DVTSLDWSNGTLATGSDGPFARIWTKDGNLASTLQHKGPFPALKNKGNFILLSAGV 286
DB 239 DVTSLDWSNGTLATGSDGPFARIWTKDGNLASTLQHKGPFPALKNKGNFILLSAGV 298
QY 287 DKTTIWDATGKAQFPFHSAPALVDWQSN--TFASCSTDMCTHVCVKLGODRIKTF 345
DB 299 DKTTIWDATGKAQFPFHSAPALVDWQ--NTTFASCSTDMCTHVCVKLGCDRPFKTF 357
QY 346 QGHTNEVNAIKWDPGTGNLLASCDDMTLKIWSKQDNCVHDLQAHNKEIYTIKWSPTGP 405
DB 358 QGHTNEVNAIKWDPGSMLLASCDDMTLKIWSKQEVCIHDLQAHNKEIYTIKWSPTGPA 417
QY 406 TNNPNANMLASAFDSTVRLVDVDRGICHTLTKEHPYVAFSPDGRYLASGSFDKC 465
DB 418 TSNPNSNMLASAFDSTVRLVDIERGVCTHTLTKEHPYVAFSPDGRYLASGSFDKC 477
QY 466 VHTWNTOTGALVHSYRGTTGIFFEVWNAAGKVGASASDGSVCVLDLRK 514
DB 478 VHTWNTOSGNLVHSYRGTTGIFFEVWNAAGKVGASASDGSVCVLDLRK 526

RESULT 7
TBLY HUMAN
ID TBLY HUMAN STANDARD; PRT; 522 AA.
AC Q9EQ87;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein Tbl1y (Transducin beta-like 1Y protein)
DE (Transducin-beta-like 1, Y-linked).
GN Name=TBLY; Synonyms=TB11;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```



```
SQ SEQUENCE 522 AA; 5668 MW; 4E020216422442D8 CRC64;
Query Match 88.5%; Score 3247.4; DB 1; Length 522;
Best Local Similarity 84.4%; Pred. No. 1.1e-56;
Matches 443; Conservative 36; Mismatches 30; Indels 16; Gaps 10;

QY 1 MSISDEVNPLVRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISIIKGLQYV 60
D 1 MSISDEVNPLVRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISIIKGLQYV 60
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVPVDVOTQOAYRDKLAQOHA-AAA--A-AAA- 115
D 61 EAEISINKDGTVDSPRIESLSLIVAVIPDVQMRQAFGEKUTQOQASAAATEASAMAK 120
QY 116 AAT-----NQQSAKNGENTANGEANGHTIANNHTDM-MEYVDGVEIPSNKAVLGRH 168
D 121 AATMTAAISQQPNKREATVNGEANGAHEI-NNHS-KPEWIDGDVEIPPNKATVLRGH 178
QY 169 ESEVFTCAWNPVSDLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIRREGQDVPSNKDV 228
D 179 ESEVFTCAWNPVSDLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIRREGQDVPSNKDV 238
QY 229 TSLDWNSECTLLATGSDVGFARITWKDGNLASTLGQHKGPFIKWKKNKGNFILSAGVDK 288
D 239 TSLDWNSECTLLATGSDVGFARITWKDGNLASTLGQHKGPFIKWKKNKGNFILSAGVDK 298
QY 289 TTIIDAHTEAKQPPHSPALPDVDWQSN--TFASCSTDMCIHVCKLGODRPIKTFOG 347
D 299 TTIIDAHTEAKQPPHSPALPDVDWQ--NNMTFASCSTDMCIHVCKLGODRPIKTFOG 357
QY 348 HTNEVNAIKWDPGNLLASCDSDMTLKWSMKQDNCVHDLQAHNKEIYTIKNSPTGPGTN 407
D 358 HTNEVNAIKWDPGNLLASCDSDMTLKWSMKQDNCVHDLQAHNKEIYTIKNSPTGPGTN 417
QY 408 NPANMLASASPDSTVRLWDVDRGICIHITLTKHQPVSVAFPDGRVYASFPKCVH 467
D 418 NPANMLASASPDSTVRLWDVDRGICIHITLTKHQPVSVAFPDGRVYASFPKCVH 477
QY 468 IWTQTGALVHSYRGTTGFEVNCNAAAGDKVGASDGSVCVLDL 512
D 478 IWTQTGALVHSYRGTTGFEVNCNAAAGDKVGASDGSVCVLDL 522

RESULT 8
Q70371 PRELIMINARY; PRT; 511 AA.
AC Q70371
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE AgCP11563 (Fragment).
GN Name=agCG45469; ORFNames=ENSANG00000019755;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 8 WD repeats.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008964; EAA12470.1; -.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.

DR PROSITE; PS0082; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT NON TER 1
SQ SEQUENCE 511 AA; 55631 MW; 7C4C2670316A8580 CRC64;

Query Match 80.1%; Score 2937.5; DB 2; Length 511;
Best Local Similarity 75.6%; Pred. No. 4e-50;
Matches 405; Conservative 46; Mismatches 30; Indels 55; Gaps 24;

QY 1 MSISDEVNPLVRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISIIKGLQYV 60
D 8 MSISDEVNPLVRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISIIKGLQYV 67
QY 61 EAEVSINEDGT--LFDGRPIESLSLIDAVPVDVOTQOAYRDKLAQOHA-AAAAA- 117
D 68 EAEISIGEDGTQORL--G---ESLSLIDAVMEVVARQ---NM---Q----- 104
QY 118 TNOQGS-KNGE--NTANG--BENGAIH--TIANNHT-----DMVEVDGVEIP-SNK 161
D 105 -NQKQATKT-BPPET-NGTTEB-PAPPPPTTTTATT-TPAPVETMEVDQSIEIPAS-K 158
QY 162 AVVLRGHESEVFICAWNPVS-DLLVSGSGDSTARIWNLSENSTSGPTQLVLRHCIRREGQ 220
D 159 AVVLRGHESEVFICAWNP-STDLLASGSGDSTARIWMSDN-PANPNQLVLRHCIRREGQ 216
QY 221 DVPSNKDVTSLDWSNSETLLATGSDVGFARIW-TKDGNLASTLGQHKGPFIKWKKNKGN 279
D 217 DVPSNKDVTSLDWSNSETLLATGSDVGFARIW-TKDGNLASTLGQHKGPFIKWKKNKGN 275
QY 280 FILSAGVDKTTIWDATGCE-AKQPPHSPALPDVDWQSNNTFASCSTDMCIHVCKLGQ 338
D 276 YILSAGVDKTTIWDATGCT-QQFSFSPALPDVDWQSNQSFASCSTDDQCIHVCKLG 334
QY 339 DRPIKTFOGHTNEVNAIKWDPGNLLASCDSDMTLKWSMKQDNCVHDLQAHNKEIYTIK 398
D 335 DRPIKSFQGHTEVNAIKWDPGNLLASCDSDMTLKWSMKQDNCVHDLQAHNKEIYTIK 394
QY 399 WSPGTGCTNNPNNMLASASPDSTVRLWDVDRGICIHITLTKHQPVSVAFPDGRVYLA 458
D 395 WSPGTGCTNNPNNMLASASPDSTVRLWDVDRGICIHITLTKHQPVSVAFPDGRVYLA 454
QY 459 SGSPDKCVHWTQTGALVHSYRGTTGFEVNCNAAAGDKVGASDGSVCVLDLRK 514
D 455 SGSPDKCVHWTQTGALVHSYRGTTGFEVNCNAAAGDKVGASDGSVCVLDLRK 510

RESULT 9
EBI DROME
ID EBI DROME STANDARD; PRT; 700 AA.
AC Q95RJ9; Q9XZK1;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein ebi.
GN Name=ebi; ORFNames=CG4063;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.; FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND MUTANT EBI4.
RX MEDLINE=99234084; PubMed=10215623;
RA Dong X., Tsuda L., Zavitz K.H., Lin M., Li S., Carthew R.W.,
RA Zipursky S.L.;
RT "ebi regulates epidermal growth factor receptor signaling pathways in
RT Drosophila.";
RL Genes Dev. 13:954-965 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=Berkeley;
```

RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fogle C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laško P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkeley; TISSUE=Embryo;
 RC MEDLINE=22426066; PubMed=12537569;
 RX Scapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A *Drosophila* full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 RN [4]
 RP FUNCTION, INTERACTIONS WITH PHYL AND SINA, AND MUTANTS EBIC1 AND
 RP EBIC3.
 RX PubMed=11032805;
 RA Boulton S.J., Brook A., Staehling-Hampton K., Heitzler P., Dyson N.;
 RT "A role for Ebi in neuronal cell cycle control.";
 RL EMBO J. 19:5376-5386 (2000).
 RN [5]
 RP FUNCTION, AND INTERACTIONS WITH SNO; SMR AND SU(H).
 RX PubMed=12230979;
 RA Teuda L., Nagaraj R., Zipursky S.L., Banerjee U.;
 RT "An EGFR/Ebi/Sno pathway promotes delta expression by inactivating
 RT Su(H)/SMRT repression during inductive notch signaling.";
 RL Cell 110:625-637(2002).
 CC -1- FUNCTION: F-box-like component of E3 ubiquitin ligase complexes;
 CC involved in R7 photoreceptor cell differentiation, cone cell
 CC development and neuronal cell cycle control. E3 ubiquitin ligase
 CC complexes mediate ubiquitination and subsequent proteasomal
 CC degradation of target proteins. Required for specification of R7
 CC photoreceptor cell fate in the eye by participating to the
 CC ubiquitination and subsequent proteasomal degradation of trantrack
 CC (ttk), a general inhibitor of photoreceptor differentiation.
 CC Required to block the S phase entry in the peripheral nervous
 CC system and central nervous system in a process that does not
 CC involve the degradation of ttk. Involved in cone cell development

CC by preventing the transcriptional repression mediated by Su(H) on
 CC Di, probably by participating to a E3 complex that contains sno
 CC and mediates the ubiquitination and subsequent proteasomal
 CC degradation of some component of the Su(H) repressor complex.
 CC -1- SUBUNIT: Component of some E3 complex at least composed of sina,
 CC ebi and phyl, required for the degradation of ttk. Probably forms
 CC a E3 complex with sno, required for the degradation of some
 CC component of the Su(H) repressor complex. Interacts with sno and
 CC Su(H) and Smr.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Widely expressed both in embryos and larvae.
 CC -1- DOMAIN: The F-box-like domain is related to the F-box domain, and
 CC apparently displays the same function as component of ubiquitin E3
 CC ligase complexes.
 CC -1- MISCELLANEOUS: 'Ebi' means 'shrimp' in Japanese.
 CC -1- SIMILARITY: Belongs to the WD-repeat EBI family.
 CC -1- SIMILARITY: Contains 1 F-box-like domain.
 CC -1- SIMILARITY: Contains 1 Lish domain.
 CC -1- SIMILARITY: Contains 8 WD repeats.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to a
 CC frameshift in position 484.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF146345; AAD35017.1; --
 CC EMBL; AB003589; AAF51501.1; --
 CC EMBL; AY061326; AAL28874.1; ALT_FRAME.
 CC HSSP; P16649; 1ERJ.
 CC FlyBase; FBgn0023444; ebi.
 CC GO; GO:0000074; P:regulation of cell cycle; IMP.
 CC InterPro; IPR006594; Lish.
 CC InterPro; IPR001680; WD40.
 CC Pfam; PF00400; WD40; 3.
 CC PRINTS; PR00320; GPROTEINRPT.
 CC ProDom; PD000018; WD40; 3.
 CC SMART; SM00667; Lish; 1.
 CC SMART; SM00320; WD40; 8.
 CC PROSITE; PS50896; LISH; 1.
 CC PROSITE; PS00678; WD_REPEATS_1; 3.
 CC PROSITE; PS50082; WD_REPEATS_2; 6.
 CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
 CC Cell cycle; Nuclear protein; Repeat; Ub1 conjugation pathway; Vision;
 KW WD repeat.
 FT DOMAIN 4 36 Lish.
 FT REPEAT 41 84 F-box-like.
 FT REPEAT 353 392 WD 1.
 FT REPEAT 408 447 WD 2.
 FT REPEAT 449 488 WD 3.
 FT REPEAT 491 531 WD 4.
 FT REPEAT 532 571 WD 5.
 FT REPEAT 574 622 WD 6.
 FT REPEAT 625 664 WD 7.
 FT REPEAT 666 700 WD 8.
 FT DOMAIN 206 323 Ser-rich.
 FT MUTAGEN 16 16 L->Q; In ebiC1, induces ectopic S phases
 FT within the peripheral and central nervous
 FT system.
 FT C->Y; In ebi4, induces defects in R7 cell
 FT development in the eye.
 FT S->L; In ebiC3, induces ectopic S phases
 FT within the peripheral and central nervous
 FT system.
 SQ SEQUENCE 700 AA; 72387 MW; 28C6D80D7B79FB7 CRC64;
 Query Match 79.5%; Score 2916.9; DB 1; Length 700;
 Best Local Similarity 57.3%; Pred. No. 1.9e-49;
 Matches 411; Conservative 51; Mismatches 34; Indels 221; Gaps 61;

```
QY 1 MSSSDEVNLFVRYLQESGFSASFTFGIESHISQSNINGALVPPAALISIIQKGLQVY 60
DB 1 MSFSSDEVNLFVRYLQESGFLSHAVVFGIESHISQSNINGALVPPAALITIIQKGLTYT 60
QY 61 EAEVSNEDGTDF-GRPIESLIDAVMPDV-----VOTRQO-----AYRD---KLA-- 104
DB 61 EVMSVGEDG---EVARPIEGLSLIDAVMPDVPLKPIVKT-BPGKPGAV-DSSAP-AGG 114
QY 105 -Q-----OH--A-A-A-A-----A-----AA-A-A-----A 115
DB 115 NQNNNAKPEIKIEPGTGVAGSAGNKIAGTTGTSTPTDOSASEVDSSGNAANNAGGTVA 174
QY 116 ---A---A-T---N-----Q-----QSAK--N-G-BN---T-- 130
DB 175 GNGAGGAGQASTGSGNSTSTPAGGDLAAPCASQKQNSNEAGSSGSSGNAAGNATSTDD 234
QY 131 -A-----NG-----EE-----N-----G-AH-----TI- 141
DB 235 AASSTSTNGNSTSSSSVEQPTSLTGPAGGTVSTSNPDAAAGSAGSTATGSKAPSGAVTIR 294
QY 142 --A--NN-----H-----T-DMMEV--DGD--VEIP-SNKAVALVR 166
DB 295 VQAQNNVQSGSSNAGSAPSGTISSTSGAGTTPAAL-VPMDIDENIEIPES-KARVLR 352
QY 167 GHESEVFIKAWNPVS-DLLVSGSGDSTARIWNLS-NSTSGPTQLVLRHCIREGGODVPS 224
DB 353 GHESEVFIKAWNP-SRDLASGSGDSTARIWMSDANTNS--NQLVLRHCIOKGAEVPS 409
QY 225 NKDVTSLDMSNBTLLATGYDGFARIW-TKQGNLASTLQHGKPIFALKKWK--GNFIL 282
DB 410 NKDVTSLDMSNCDGSLATGYDYARIWKT-DGRLASTLQHGKPIFALKWN-KCGNYIL 467
QY 283 SAGVDKTTIWDATGE-AKQPPFHSAPALVDVQSNNTFASCSTD-MCIHVCKLG--Q 338
DB 468 SAGVDKTTIWDASTQOCT-QQAFHSAAPALVDVQTNQAFASCSTDQR-IHVCRLGVNE 525
QY 339 DRIKTFQGHTEVNAIKWDPCTGNLLASCDDMTLKIWSMKOD-NCVHDLQAHNKIYTI 397
DB 526 --PIKTFKGTNEVNAIKWCPQQLASCSDDMTLKIWSNDRCC-HDLQAHNKIYTI 582
QY 398 KWSPTGFGTNNPNANMLASASFDSTVRLWVDVDRGICHTLTKHQBPFVYSAFSPDGRYL 457
DB 583 KWSPTGFGTNNPNNTLILASASFDSTVRLWVDVDRGSCIIHTLTKHTEPVSVAFSPDGKHL 642
QY 458 ASGSFDPKCVHINWOTGALVHSYRGTGTFEVCNNAAGDKVGASADGSGVCVLDLRK 514
DB 643 ASGSFDPKCVHINWOTGQLVHSYKGTGTFEVCNWSKGTKVGSASDGSVFVLDLRK 699

RESULT 10
Q96RW7 PRELIMINARY; PRT; 5636 AA.
AC Q96RW7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hemiscetin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Trent J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156100; AAK68690.1; --
DR HSSP; P07996; 1LSL.
DR GO; GO:0005824; F:ATP binding; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005021; F:vascular endothelial growth factor receptor. .; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
```

```
DR InterPro; IPR000875; Cecropin.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006605; G2F.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR000169; Pept_cys_acsite.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR009134; VEGFR.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF07645; EGF_CA; 8.
DR Pfam; PF07474; G2F; 1.
DR Pfam; PF00047; Ig; 44.
DR Pfam; PF00090; TSP_1; 6.
DR PRINTS; PR01832; VEGRECEPTOR.
DR SMART; SM00179; EGF_CA; 7.
DR SMART; SM00408; IGC2; 43.
DR SMART; SM00209; TSP1; 6.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00268; CECROPIN; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS00026; EGF_3; 5.
DR PROSITE; PS01187; EGF_CA; 8.
DR PROSITE; PS00835; IG_LIKE; 44.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
DR PROSITE; PS00092; TSP1; 6.
KW EGF-like domain.
SQ SEQUENCE 5636 AA; 613673 MW; F000B319CED7B52C CRC64;

Query Match 71.6%; Score 2627.8; DB 2; Length 5636;
Best Local Similarity 9.6%; Pred.No.1.3e-41;
Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;

QY 1 M-S-----I-----SS---D-----EV----- 8
DB 1 MISWEVHTVFLFALLYSLAQDASQSEIRABEFPPEGASTLAFVFDVTGSMYDDLQVI 60
QY 9 -----NF-LV----- 12
DB 61 EGASKILETSLRKPRLFPFALVPPHDPPEIGPVTITDPKPKQYBELRELYVGGDCPE 120
QY 13 -----Y-----R---Y-----LO----- 17
DB 121 MSIGAIIKALEISLPGSFYVFTDARKSKDYRLTHEVLQIIQQKQSVFVLTDGCDRTH 180
QY 18 -----E---SG--F-----S--H--S-----A---F 26
DB 181 IGVKYVEEIASTSSGOVFFHLDKKQVNEVLKWEAAQASKVHLLSTDHLEQAVNTWRIPF 240
QY 27 -----T-----FG-----I-----E----- 31
DB 241 DPSLKEVTVLSGSPSPMIBIRNPLGKLKKGFGLHELLNHNSAKVNVNKEPEAGMWTVK 300
QY 32 ---S--H-----I-----S---Q----- 36
DB 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSRPVGPIPTYVLLNTSGISTPARID 360
QY 37 -----S-----NI-----N-----G----- 41
DB 361 LLELLISGSSSLKTPVKYYPHRKPYGIWNIISDFVPPNEAFFLKVTGYDKDYLQFVSS 420
QY 42 -----A-----L---VP-----P-----A 47
DB 421 VSFSSIVDPAPKVTMPEKTPGYVLPQGPICSDLSLPTLSFVRNGVTGLGVQYLKESA 480
QY 48 -----A---L-----I-----S---IIQ-----KG----- 56
DB 481 SVSLDIATKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPPIQVNNVTVTGERAVLT 540
QY 57 -----L---Q-----Y--V--E----- 61
```

Db 541 CLIIAIVDYNLTQWRNDRVRLAEPARIRTLANLSLELKSVDKENDAGEYHCWVSSEGGSS 600
QY 62 -A-----EVS-----NE-----68
Db 601 AASVFLTVQBPVKVTVMPKNQSFSGSEVIMSGATGPKPKIAWTVNDMFIVGSRYRM 660
QY 69 --DGLF-----G-----R-----P-----I--78
Db 661 TSDGTLFIKNAAPKADAGIYCLASNSAGTDKQNSTLYRIEAPKLMVQVSELLVALGDITV 720
QY 79 -E--S-----L--S--LI-----DA-----V-----87
Db 721 MECKTSGIPPPQKWKFGDLELSPSTLIIDPLGLLKIQTQDLDAQDYTCVAINEAGR 780
QY 88 -----M-----P-DV-----VQ-----T-----94
Db 781 ATGKITLDVGSPPVFIQEPADVSMIEGNSVTLPCYVQGYPEPTIKWRRLDNMFISRPFS 840
QY 95 -----R-----OH-----A-----108
Db 841 VSSISQRTGALFILNLWASDKGTIYCEAENQFGKIQSETTVTGLVAPLIGISPSVAN 900
QY 96 ---Q-----A-----Y--R-D-KL-----AQ-----105
Db 901 VIEGQQLTLPCTLLAGNPPIPERWIKNSAMLLQNPIYTVRSDSLHIERVOLQDGGEYTC 960
QY 106 -----OH-----A-----108
Db 961 VASNVAGTNNKTSVVVHVLTQHQQILSTIEGIPVILPCKASGNPKPSVIWKKGEL 1020
QY 109 -----A--A-A-----A--A-A-----114
Db 1021 ISTSSAKFSAGAGSLYVSPGEESGEYVCTANTAGYAKRKVQLTVYVPRVFGDLRG 1080
QY 115 -----A-----A-----116
Db 1081 LSQDKPVEISVLAGEEVLTPCEVKSLLPPPIITWAKETQLISPFSPRHTFLPSGSMKITET 1140
QY 117 -----ATN-----Q-----Q-GSA--K-----N-----126
Db 1141 RTSDSGMYLCVATNIAGNVQAVKLVNHVPPKIQRGPKHLKLVQVGQRPDI PCNAQGTPLP 1200
QY 127 -----GE--N-----T--A--N--G-----133
Db 1201 VITWSKGGSTMLVDGHHVNDPDTLSIDQATSDAGIYTCVATNIAGTDETEILHVQE 1260
QY 134 -----E--N-----GA-----H-----139
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEFPCKAGTPKPTIKWLHNGREL TGREP GISILEEG 1320
QY 140 T--IA-----N-----N--HT-----DM-----148
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKQEQVNSVLLNQLTNL 1380
QY 149 -----M-----E-----150
Db 1381 FCEVEGTPSPIIMWKNDVQVTSSTIQTVNNKILKLFATPAGEDAGRYCKAINIAGTS 1440
QY 151 -----VD-----G-----DV--E-----I-----P--158
Db 1441 QKYFNIDVLPPTIIGTNFPKEVSVLNRDVALEQCVKGTFFPDDIHWFKDGKPLFLGDPN 1500
QY 159 -----SN--K-A-----V-----VL 165
Db 1501 VELLDRGQVLHKNARNDKRGYQCTVSAAGKQAKDKLTIYIPPSIKGNVTTDISVL 1560
QY 166 -----RG-----H-----168
Db 1561 INSLIKLECTRGLPMPAITWKDGPIMSSQALYIDKQOYLHI PRAQVSDSATYTCVH 1620
QY 169 -----E-S-----E-----VFI-----C--A-----W-----177

Db 1621 ANVAGTAERKSFHVDVYVPPMIEGNLATPLNKQVVIASHSLTECNAAAGNPSPIITLWKDGV 1680
QY 178 -----N-----P-V--SD--182
Db 1681 PVKANDNFRIAGGKKLEIMSQAQIEDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740
QY 183 -----LL-----VSGS-----G-----D-----S--191
Db 1741 TSFYIWMNNLLELDCHVTGSPPTIMLWKDGLIDERDGFKILLNGRKLVIQAQVSNT 1800
QY 192 -----TA-----R-----I--W 196
Db 1801 GLYRCMAANTAGHKKEFEVTVHPPTIKSSGLSERVVVKYKVALQCIANGIPNPSITW 1860
QY 197 -----NL-----S-----E-----NS--T-----S 204
Db 1861 LKDDQVNTAQGNLKTQSSGRVLQIAKTLLEDAGRVTCTVATNAAGETQHQIHLHVHEPPS 1920
QY 205 -----G-----PTQL-----V-----L-R-----212
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPIWYKDNCLLSGSTSTMTFLNRQIIDIES 1980
QY 213 -----H-----C-----I--215
Db 1981 AQISDAGIYKCAINSAGATFYSLSQVHVAPSISGSNNMVAVVVWNPVRLECEARGIPA 2040
QY 216 -----R-----E--G-----G--Q--D--221
Db 2041 PSUTLWKDGPVSSFSFNGLVLSGGRIALATSTQISDTGRYTCVAVNAAGEKORDILRV 2100
QY 222 -VPSN-----KD-----VT--S-L--231
Db 2101 YPPNIMGEOQVSVLSQAVELLQSDAI PPPTLTWLKDGHPLKPKGLSISENSVLK 2160
QY 232 -----D-----W--N--S-----EG--TLL--240
Db 2161 IEDAQVQDGRYTCVATNVAGTEKKNVYNIWPPNIGGSDLTQLTVEGNLISLCE 2220
QY 241 AT-----GS--Y-D-----G-----F-----A--249
Db 2221 SSGIPPNNLIWKKGSFVLTDMSGRVRIILSGGQLOQISAEKSDAALYSCVANSVAGTAK 2280
QY 250 -----R-----I-----WTKDG--256
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSISLECEVQGPPTPTVWMDGHPILIRAKGV 2340
QY 257 -----NL-AS-----T--L-----GOH--265
Db 2341 EILDEGHILQKNIHVSDTGRYVCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400
QY 266 -K-----G-PI-----FALK--W-----274
Db 2401 ERKNSVLTCEASGIPLPSTTWFF--KDGWPVLSVNSVRILSGGRMLRMQTMEDAGQYTC 2458
QY 275 -----N-----K-K-----GN-----279
Db 2459 VVENAAGEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNVPVPIWTKDQ 2518
QY 280 -----F--I-----L--S--AGV--286
Db 2519 QPLQDEDAHHIISGGRFLOITNVQVPHTRGYTCCLASSPAGHKSRFSLSNVFVSPTIAGVG 2578
QY 287 -D--K--T-----T--I-----I--W-----293
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPFATITWFKDGTPLSNRNRILPGRRTLQILNAQ 2638
QY 294 -D-----A--H-----TG-----E 299
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVYIPIINKGDLWGLSPGKVKVKNVNTLTLECE 2698
QY 300 A-----K-----Q--Q-----303
Db 2699 AVAIPASLASWYKDGQPLKSDDDHVNIAANGHTLQI KEAQISDTGRYTCVANSIAGEDLD 2758

QY 304 F---P-F---H- 307
Db 2759 FDVNIQVPPSFQKLEIGNMLDTGRNGEAKDVIINNPISLYCETNAAPPTLTWYKDGHP 2818
QY 308 --SA---P-----A-----LDV-----D 315
Db 2819 LTSSDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878
QY 316 -----WO-----SN----- 319
Db 2879 LPEVTVLVNKSALIECLSSGSPAPRNSWOKDGOPLLEDHDKFLNSGRILQILNTQITD 2938
QY 320 -----NT-----FAS--C-----S 326
Db 2939 IGRYVCAENTAGSACKYFNLNVHVPSPVIGPKSENLTVVVNNFISLTCEVSGFPDLS 2998
QY 327 -----T-----D-----MCI----- 331
Db 2999 WLKNXOPIKLNTLIVPGGRTLIIRAKVSDGGEVTCIAINXAGESKKFSLTVVPPS 3058
QY 332 --H-----V--C-----K-----L-GO----- 338
Db 3059 IKDHSLSLVNVRECTSVSLECESNAVPPVITWYKNGRMITESTHVEILADGQMLHI 3118
QY 339 -----D-----R-----P-IK----- 343
Db 3119 KKAESDVTGOYVCRAINAGRDDKNFHLNVVYVPPSIEGPEREVIVETISNPVTLTCDATG 3178
QY 344 -----T-----F-----O--G-----H-----T----- 349
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGSKLQIARSQSDSGNYTCIASNMEGKAQKY 3238
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360
Db 3239 FLISQVPPSVAGAEIPSDVSVLGENVELVCNANGIPTPLIOWLKDGKPIASGETERIRV 3298
QY 361 -----G--NLL-A--S-----C--S-----D-----D-----MTL-- 373
Db 3299 SANGSTILNIYGALTSDTGKYTCVATPAGEDIRFNLNVVVTPIRGNKDEAKLMTYVD 3358
QY 374 -----K-----I-W-----SM----- 378
Db 3359 TSINIECRXTGTPPPQINMLKNGLPLPLSSHIRLLAAGVIRIVRAQVSDVAVVTCVASN 3418
QY 379 -----K-----Q-----DN-----CV----- 384
Db 3419 RAGVDNKNHYNLQVFPAPNMDNSMGTBEITVLKGSSTSMACITDGTAPSNMWLRDGOPLG 3478
QY 385 -----H-----D--L-----QA-----H-N-----KE 393
Db 3479 LDAHLSVTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKLVLEPPHINGSEHEE 3538
QY 394 I-----YT-----I--K--W----- 399
Db 3539 ISVIVNNPLELTCIASGIPAPKMTWMMKGRPLQTDQVOTLGGGEVLRISTAQVEDTGRV 3598
QY 400 -----SPTG-----P--GT-----N----- 407
Db 3599 TCLASSPAGDDOKEYLVRVHVPNIAGTDEPRDITVLRNRQVLTCKSDAVPPPVITWLR 3658
QY 408 N-----P-----N-----ANLM-LAS-A--S--F----- 420
Db 3659 NGERLOATPRVRLISGRVLIQINNADLGDNTANTCVASNIAGTKTREFILTNNVPPNIKG 3718
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433
Db 3719 GPQSLVILLNKSVTLECIAGVPTPRIITWRKOGAVLGNHARYSILENGFLHIQSAHVT 3778
QY 434 -----C-----I-----H-----TL-----T----- 439
Db 3779 TGRYLCWATNAGTDRRRIDLVHVPVPSIAPGPTNMTVIVNVTTLACEATGIPKPSINW 3838

QY 440 -K--H-----Q-E-----P-V-----Y----- 446
Db 3839 RKNGLHLLNVQDNQNSYRLLSSGLSVIISPSVODTATVECTVTNGAGDDKRTVDLTQVQPP 3998
QY 447 SVA-----F-S-----P-D----- 453
Db 3899 STADEPTDFLVTKHAPAVITCTASGVPPFSIHWTKNGIRLLPRGDGYRLLSSGAIEILAT 3958
QY 454 -----CRY-----L-----ASG--S-- 461
Db 3959 QLNHAGRYTCVARNAGSAHRHVTLVHVEPPVIOPOPSLHVILNPNPILLPCEATGTPSP 4018
QY 462 F-----D-----KCV----- 466
Db 4019 FITWQKEGINVNTSGRHAFLPSGGIQLISRAVEDAGTVMCAQNAGTALGRIKLNQVQ 4078
QY 467 -----H-----I-W----- 469
Db 4079 PPVISPHLKEYVIADVDPITLSCADGLPPDPITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 470 -----N-----T-----Q----- 472
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIIRSTEGHYTVNENSQAILPCVADGPT 4198
QY 473 -----T-----G-----A-----L-VH 478
Db 4199 PAINMKKNVLLANLILGKYTABPYGELILENVLEDSGFYTCVANNAGEDTHTVSLTVH 4258
QY 479 -----S-----Y-----R 481
Db 4259 VLPTFTPELQVSLNKGELRLSCKATGIPKLTWTNNIIPAHFDSVNGHSELVIER 4318
QY 482 -----GT-----G--G-----IF-----E-----V--C----- 490
Db 4319 VSKEDSGTYVTAENSVGFKATGFVYVKEPPVFKGDYPSNWTIEPLGGNAILNCEVKGDP 4378
QY 491 -----WN-----A-----AGD----- 496
Db 4379 TPTIQNRKGVDEISHRIRQLNGSLAIYGTVNEDAGDTCVATNEAGVVERMSITLQ 4438
QY 497 -----K-V--G-----A-----SAS----- 503
Db 4439 SPPIITLBPVETVINAAGKIILNCOATGPQPTITWSRQHSISWDRVNLNNSLYIA 4498
QY 504 -----D-----GSV-----C-V----- 509
Db 4499 DAQKEDTSFEFCVARNLMGSLVVRVPIVQVHGGFQNSAWRACSVTCGKGIKRRSLCN 4558
QY 510 --L-----DL--R--K 514
Db 4559 QPLPANGKPCOGSDLEMCNK 4582

RESULT 11

Q8NJX1
ID Q8NJX1 PRELIMINARY; PRT: 20925 AA.
AC Q8NJX1
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Nonribosomal peptide synthetase.
GN Name=txl;
OS Trichoderma virens.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaeae; Hypocrea.
OX NCBI_TaxID=29875;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22037864; PubMed=11909873;
RA Wiest A., Grzegorski D., Xu B.W., Goulard C., Rebuffat S.,
RA Ebbolle D.J., Bodo B., Kenerley C.;
RT "Identification of peptaibols from Trichoderma virens and cloning of a
peptaibol synthetase.";

J. Biol. Chem. 277:20862-20868(2002).

[2]
SEQUENCE FROM N.A.
Wiest A.E., Grzegorski D., Xu B.-W., Zhou W., Ebbole D.J.,
Kenerley C.;
Submitted (JAN-2002) to the EMBL/GenBank/DBJP databases.
CC - SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC EMBL; AF469045; AAU78457.1; -.
DR HSSP; Q30409; IDNY.
DR GO; GO:0048037; F:cofactor binding; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR010071; AA_adenyl_dom.
DR InterPro; IPR009081; ACP_like.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000873; AMP_bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR000794; Ketoacyl synth.
DR InterPro; IPR006162; Ppatne_S.
DR InterPro; IPR006163; Pp_bind.
DR InterPro; IPR010080; Thioester_redct.
DR Pfam; PF00698; Acyl_transf_1; 1.
DR Pfam; PF00501; AMP-Binding; 18.
DR Pfam; PF00668; Condensation; 27.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF00550; PP-binding; 19.
DR PRINTS; PR00154; AMPBINDING.
DR TIGRFAMS; TIGR01733; AA-adenyl-dom; 17.
DR TIGRFAMS; TIGR01746; Thioester-redct; 1.
DR PROSITE; PS00075; ACP DOMAIN; 19.
DR PROSITE; PS00455; AMP-BINDING; 18.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN 11.
DR QUERY 20925 AA; 2320297 MW; C6603A269D56F11 CRC64;

Query Match 71.5%; Score 2622.8; DB 2; Length 20925;
Best Local Similarity 8.2%; Pred. No. 1.8e-40;
Matches 477; Conservative 30; Mismatches 5; Indels 5312; Gaps 391;

QY 1 M-----S-ISS-----D-----E-----V----- 8
DB 8843 MISVEYQLQNISKISHAMEACDFGSLIIIOPIEQLAYRGDDALLVAESHTNETSAL 8902
QY 9 -N-F-----LVY-----R-----Y--IQE----- 18
DB 8903 QNYFSYPLVIQCHLYDNTINLMILIYDSQALSSESMELSHFDNAVQELASKDNLLIGSI 8962
QY 19 S-G-F-----S-----H-----S-A-----FTF----- 28
DB 8963 SLAGRWDFEQALKLNSDSPDAVERCFHEMLDETALVRGDSLAIAGWOKSFYTREMAETT 9022
QY 29 -----G--I--E-----SH-----I-- 34
DB 9023 RIAHLHYTDYGKVGDIIHVCFEKSAFWFIATLAINKAGAAMSTLDPSHPTRYQKIINQ 9082
QY 35 --SQ-----SN----- 38
DB 9083 TGSQTALSSPVNSRCASLENNVIELTSDELSDKLQAQNVSWSTRPAVNVTFRDAAYVLT 9142
QY 39 -----I-----N-----VP-----GA-42
DB 9143 SGSTGVPKGVIEHGSLCTSQSLSKALDFNEEFRLVFQSSYSPFDIILFEIGSTFFLTGAC 9202
QY 43 -----L-----P----- 46
DB 9203 LFVPSWDEQMNELVEYIRKHQITFWVLTPTLTARTRPEDVPSVDMLVAGEAPTRDLDI 9262
QY 47 -----A-----A-L-----I-S----- 51
DB 9263 WFGKURLANGWGPTCSVIACLHOWTSVDESFPKVGIRPGIGSCWIVDPDATCMAPLGTV 9322
QY 52 --I-IQ-----K-----GL-Q-----Y-----V-EA--E 63

Db 10403 DPENPHRLAPTALGEVIOGPTILREYLSVDORTEAAVITKSLPDWAPFREQSGSRFNK 10462
Qy 162 -----AV-V-L----- 165
Db 10463 SGDLGVNPDGTIEFSSKDTQVKIRGLRVELGEVEHVAQVALDGVHQIAVDVPKGDNGT 10522
Qy 166 --RG--HE--S--E-----V-----FI-C-- 175
Db 10523 QPRGYFSDESROIHEAHPGSPQAIDKQLARLTVAIGKLNIALPRYMIPTLFPCKY 10582
Qy 176 -----A-----W-----NP-- 179
Db 10583 MPSITSKLDRNELRRRTLLLTQSELAWFSYRNCNKRTPETPMESLIQRIWSEILHPLDS 10642
Qy 180 -----VS-----D-L-----L----- 184
Db 10643 IGRDPSFLGGDSITAIHLGSISSROGISITAKVDFFDPLWDIASKARELTEAQOQS 10702
Qy 185 --VS--GS-----G----- 189
Db 10703 PDISPFSLIGETRELAAGSVKQKLKLTSTQIEDAYPCTKIPGSIYAKYVYRLPHOV 10762
Qy 190 D-----S--T-A-----RI-----W-NL-S-----E-N-----S 202
Db 10763 DIGRFKSSMEHTVALHPILRLTRIIMIKOTCVQLVVKDSVTWENLDSGDLDEAVNATHSYS 10822
Qy 203 -T-----SG-----P----- 206
Db 10823 MTYGSPLSRYSIHESKGDYFLWTAHSHIDGWSVPVINTLYQAYKGLGPKPAYS 10882
Qy 207 -T-----OL----- 209
Db 10883 FIKYTMEHOKAAGDYWEQLONAKRASPPKTKLAAKSIWTKRTRWMTSLKFSASSNE 10942
Qy 210 -----H-----V-L-----R----- 212
Db 10943 IATKASVIRAAMAVILARYCESDDVCFGATISGRQASVPLGLEMAGPATAVTPVRVQLDN 11002
Qy 213 -----H-----C-----I-R----- 216
Db 11003 DQEISKFLQNIQSOAHMPYEQYGLQSTAKLGNARDACDFTSLVVIQIQRQBSVQS 11062
Qy 217 -----E-GG-----Q-D-----V-----P-- 223
Db 11063 DDALLIPVETELEDILQSYNYPLVLQGHYNDYADLVLYDSTVSVSEPMQMTALCHFFNN 11122
Qy 217 -----E-GG-----Q-D-----V-----P-- 223
Db 11123 IVOQLLAEDGGKLGDISIASSFDLEAQRSGDGPQIIDDCHLIIRERQAKQRPNRP 11182
Qy 224 -----S-----NK----- 226
Db 11183 DAWDGFTYSELDRNTANRLAHLVHDYAVKVGDIHVCFEKSRYVFAILAVNKAAGAAWA 11242
Qy 227 --D-----V-----T-----S-LD--W-----NSE- 236
Db 11243 PFDPAHPQRLKAVASOTGAKLALASTANTRLCEQVQVDCVVESSVTLDKNLWTYDNIK 11302
Qy 237 G-----T-----L-----A-----T----- 242
Db 11303 GPDINVTMDAAVILFTSGSTGVPGKIVMKGALCTNOAALSGWLGFHTVRMLQPSFV 11362
Qy 243 -----G-----S-----Y-----D----- 246
Db 11363 FVSVGEIWOALMNGACVCPSEHMLNSLDSFVRDNFTWYALTPTSFTRLKPKDFPSL 11422
Qy 247 -----G-----F-----A-----R----- 250
Db 11423 KULLLAGEPTQDVLDTWGLPNTRFINAWGPAETCVYNTLYEWSQNTSPKLGRAVCA 11482
Qy 251 -IW----- 252
Db 11483 YIWWVDVENPQRLAPTGLGEIIVQGPPLLKEYLADPEKTAATAVTPELWAPRROSTTW 11542

Qy 253 -----T-----KD----- 255
Db 11543 NRYRTGDLGYDHDGMLHFASRKDTQVKIRGLRVELGEVEHRIQSSLDGVRQVAVDVFK 11602
Qy 256 --G-NL--AS--T-----LG-- 263
Db 11603 TEKANLVAYLCTSDTKTPDQNTDFEGKDVFASIDTELQINLGKMLTELNSLPTYMIP 11662
Qy 264 -----Q----- 264
Db 11663 TMFPCAYMPFITSSKLDVRKRLRLFAELSDQOLEAYSLLVNKEQAPETEMEIRLOKLWA 11722
Qy 265 -----H-----K-G----- 267
Db 11723 EVLDLPEASIGHDNFMRIGGDSIAAIRLWSMARDAGISLTVNDIPDARLISVATKAID 11782
Qy 268 -----PI-----FAL----- 272
Db 11783 NDESHLMAPIEPFSLTSGIEDFALPSLSDQFIEDAYPCSKLOEGLMALAIKQPGSY 11842
Qy 273 --K-----W-----N----- 275
Db 11843 IAKYIYKLSEHVDVDFKAAWESTVEIASALRTRIIQTGCGSIQVWINGDISWDEADGDL 11902
Qy 276 --KKG--N--F-----I-LS-----AGV 286
Db 11903 QSCLAQSOTLEMGYTRLCRYTLIRRDNDTYFLWNHHAVIDGLSTQNLGTLFNIYSGV 11962
Qy 287 -----DK----- 288
Db 11963 DVLPLPPNRFIYQMLQLEEDTANTYWKQLHNAQRTIPPNDVARDKPAATQMLQSSIE 12022
Qy 289 --T-TI--I--W----- 293
Db 12023 LPHGLDNSTVTIATVIRAAWAILVARYCSDDDVTFTTISGRQAPIPEIMGVGTIATV 12082
Qy 294 -----DA----- 295
Db 12083 PVRVLNRQOLISDFLEGVQRAVEMIAFEQYGLQNIAKLGDDARDACDFSSLLVQPIQ 12142
Qy 296 H--T-----GE--A-K-Q-----Q-----F----- 304
Db 12143 HLAGTKDDSLVDASIGEDNAVEAJQNYFSYPLVIAQHLHYDDHIFILYDSSIMLEPQ 12202
Qy 305 -----P-----F----- 306
Db 12203 MVALSYHLOHVMQLTNTTTPATLETVSVSSYDVERALAINAEIPEVIDTCHEMFENRV 12262
Qy 307 H-S-APA-----L-D--V-D----- 315
Db 12263 HLNPLAPAAWADGNTYAEALNSAANKLAHYLIKSYSVKLNLDLVHVCDFKSAWYIVSIIA 12322
Qy 316 --W-----Q-----S-N-----N----- 320
Db 12323 INKAGATVPLDPFPHPEQRLRSIVNQTKTALASPNGITLCSALVDNVVVESSALDSML 12382
Qy 321 -----T-----PAS-----C----- 325
Db 12383 PAQDGLSPKSVSSRTAAVILFTSGSTGTPKGVQIHRSLCTNMNAIAKRVRYHLDVRI 12442
Qy 326 --ST--D--M-----C----- 330
Db 12443 LQFSAYVDFDSILEIIMSLLOGACVCPSEHIRMNGIVDFIRDMNINWLYLTPSFURTIN 12502
Qy 331 -----IH----- 332
Db 12503 PIDVNVNVLVAGGEAI PRDVFETWGVRFVINGWGPTETTTVVGSIHFEVSDESPTIG 12562
Qy 333 --V--C-----K--LG-----Q-----D-----RP 341
Db 12563 HPVGGFCWTDVNNPQLLAPTGLGEIIVQGPPLLHEYLNDPNPKTQEAAILYDLPEWAPRP 12622

QY 342 -----IKT-----F-----Q----- 346
Db 12623 DENNMGFRYKTDGLGFYNANGKIEFSRKDQTVKIRGLRVELGEIYBQVQASVEEIRQIA 12682
QY 347 -----G-----H-----TN--E-----V----- 352
Db 12683 VDVIKTDGNSLVAYLCFNDERQLNADVNGPFPPLDNLQETLAGAIGKLSVTLPYRM 12742
QY 353 -----N-----A-----I 355
Db 12743 IPTFYPCSYMPSITSGKLDKRLKQTAALSOELNKLHGVSKRAPETPMELQONI 12802
QY 356 KW -----DP----- 359
Db 12803 -WSKLLSIPSEIGRDDSFLGLGDSIMAIHLVTACREAGVSLTVKEIFDDPRLSAVASH 12861
QY 360 -----T----- 360
Db 12862 ARSMDAIDQDLPLVPFSLSDRLREMLVSDDTFQTSLLPSQVIEDAYPCSKLOEGLMAL 12921
QY 361 -----G-----N-----LL----- 364
Db 12922 SVKORGSYVAQYVYKLSGVDLVRFKASWERTIQCANLRTRIVMLDGTGVLLVDGPAB 12981
QY 365 ---A-S---C-----S-----D-----M- 371
Db 12982 WDNAFSTDQLCTLNATREMYGSLRNRYALVRDPFGNHFIVWSSHAVHDGWTLRIMN 13041
QY 372 TL----- 373
Db 13042 TLYSLYDQGLPPLLPSAFIRYTVNIDKEASRFWTEQMKNAKRATYPPMPRIESHTGV 13101
QY 374 -----K--I---W-----S-----M 378
Db 13102 SRMMNMTISPSPSVKTVNTRATILRGTWAILLARYCDTDDITFGTTVSGRQAPVPGITEM 13161
QY 379 -----K-----Q--D-----N-----C----- 383
Db 13162 PGVPVATVPIRVLEASQTVQAFSLNSIQSQTMDIAPEQFGLQNIKLTSDARDACEFSS 13221
QY 384 --V-----H-DL----- 387
Db 13222 LLVQPRSHLDLLKKGESSESLVASDARSQGLIQNYFTYPLVQGHVFEDSIELLT 13281
QY 388 -----Q-A-H-----N-----KE----- 393
Db 13282 YDSTILSEVQMKALSHQFNVANQLVNESNDPLSSITISGEWDEQAKWNVENPEILDT 13341
QY 394 -----I-----YT-----IK----- 398
Db 13342 CIHSLIEQARIKRDPAICAWGEMNYSOLNSAANOLAHILKIGIKADDLVHVCFEKS 13401
QY 399 -W--S-----PT-----G-----PGTN-----N-- 408
Db 13402 WPFVSIIAINKYGAAWVPLDPSHEQRLQVVGQTLAKFALSSP-TNAALCNKLHVNI 13460
QY 409 ---P-----N-A-----N-----L-M-----LA 416
Db 13461 EVPSLIDELSKFCDFGSPAINVPSSNAAVVLFTSGTGTGPKLVMOHGAVCTSOATA 13520
QY 417 ---S-----AS-FD-----S-T-----V--RL-W- 427
Db 13520 KRUSLTPDVRILQFAAYVFDLSIGEIVAPLIHGACVCPSEETRMNGLKEFIRDARINWA 13580
QY 428 -----DV----- 429
Db 13581 YLTPSFVTRLRPEDVPSIQLLALLAGEAVGRDILDTWFGKVLINGWGAETCVFTLHEW 13640
QY 430 ---D-----R--G-----I-C----- 434
Db 13641 SSIDESPLTIGRPVGGYCWIVEADSNKLTPIGCLGEVVLQGPFTLLREYLADPQSKETI 13700
QY 435 I-----H-----TL-----T--K----- 440

Db 13701 ITELPPWAPQVSDAHLNSRFYKSGDLCFYNFNGTLEFFYSRKDQTVKIRGLRVELGEV 13760
QY 441 H-----Q-----E-----P-----VY----- 446
Db 13761 HIRELLEGVRQVAVDLTSETGTQLVSVYICFNDDSQPSSPELKASDIYLPDLADIOARIT 13820
QY 447 -----SVA-----F----- 450
Db 13821 SMVGELSVTLPRYMIPTLPIPKYMPFTTSTKLDKRLKSLTASLGRDELHIHSLNSKK 13880
QY 451 -----S-----P-D--GR-----YLAS-----G-S----- 461
Db 13881 RVPETEMETRLQAIWSEILNLPIDISIGRDDSFLQIGDSDITAIYLVSKAREAGISLIKD 13940
QY 462 -FD-----K-C-V--HI----- 468
Db 13941 IFDDSRLLAVASKAVLSTEFQBPQGPVFPFSLNKTALVLAGEVRKFCGLNDHDIED 14000
QY 469 -----WN----- 470
Db 14001 AYPCTSLQELMALTWKQPGSVYAKYVYKLSFTVDMERFOAAWNRMTMELCGNMRTRIVLL 14060
QY 471 --T--Q-----T-----GA-----L-----V----- 477
Db 14061 NGTPIQLLLKEDNQWOLENDTLASITNSSRDLKMGYGAPLCWYGVLEENNNKYFVWSAH 14120
QY 478 HS-----Y----- 480
Db 14121 HSYDGMVNRILLTTLTYIYSTDVTLPQYSGFIKYNMELDSMALADFWREOLSGSKRA 14180
QY 481 -----R-----GT- 483
Db 14181 VFPAROTSPSSSSTQIFKSIISLEQAKOSIITKASILRAAWAIVLARYCDTDDVSGFTT 14240
QY 484 -----G----- 484
Db 14241 VSGRHAPVAGLETMPGPMIATVPVRVHLNRSATKSQFLTEIQKQAYEMVPEYQGLQNIS 14300
QY 485 -----G-----I 486
Db 14301 KLSRDARTCDFSSLVIOPPATTISEKADNTNIIYDEBOSLTDAMHINYFNPLVII 14360
QY 487 --F-----V----- 489
Db 14361 LMTFEDHILORFPFNSEVLTEARVSALSNNHIDHVVKLLAETDETLESVNLVGDWDIOHA 14420
QY 490 -----C--W-----N-AA--GD-----K-----VG 499
Db 14421 LASTRLKPSTESCTHMLIOERIKTQPNDAIVSWDGLTYKELGVLASRLAWKLQGLGVG 14480
QY 500 -----AS----- 501
Db 14481 PESLIPLCFPKSTWAVVAMVAIEMAGGAFVPLDPNAPVARLGIIBETKSSSLAVASPCQ 14540
QY 502 -----A--SD-----GS-----V----- 507
Db 14541 DTVRDIGIEVFAVDEALLLELSDPADGVKSMQAPDNASVVLFTSGTGKPKGMVQHNSL 14600
QY 508 -----C-----VLD--L-----R----- 513
Db 14601 CSSGNAYGCDLEIGPGTRIFQFSAYTFDVGVLDCVLSMRGATCIPSDHARLNDLAGAM 14660
QY 514 ---K 514
Db 14661 NVTK 14664

RESULT 12
AAP49011
ID AAP49011 PRELIMINARY; PRT; 7073 AA.
AC AAP49011;
DT 02-MAR-2004 (Tremblrel. 27, Created)

DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DE 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
OS Orifab polyprotein.
OC SARS coronavirus 2301.
OC Viruses; seRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus; Human coronavirus (strain SARS).
OX NCBI_TaxID=230471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZJ01b;
RX MEDLINE=22889812; PubMed=14527350;
RA Li L., Wang Z., Lu Y., Bao Q., Chen S., Wu N., Cheng S., Weng J.,
RA Zhang Y., Yan J., Mei L., Wang X., Zhu H., Yu Y., Zhang M., Li M.,
RA Yao J., Lu Q., Yao P., Bo X., Wo J., Wang S., Hu S.;
RT "Severe acute respiratory syndrome-associated coronavirus genotype and
RT its characterization.";
RL Chin. Med. J. 116:1288-1292(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ZJ01b;
RX PubMed=14733771;
RA Wang Z.G., Li L.J., Luo Y., Zhang J.Y., Wang M.Y., Cheng S.Y.,
RA Zhang Y.J., Wang X.M., Lu Y.Y., Wu N.P., Mei L.L., Wang Z.X.;
RT "Molecular biological analysis of genotyping and phylogeny of severe
RT acute respiratory syndrome associated coronavirus.";
RL Chin. Med. J. 117:42-48(2004).
DR EMBL: AY286320; AAP49011.4; --
KW Polyprotein.
SQ SEQUENCE 7073 AA; 790251 MW; 96E95991DACAGEA9 CRC64;

Query Match 71.4%; Score 2620.7; DB 2; Length 7073;
Best Local Similarity 8.3%; Pred. No. 2.7e-41;
Matches 469; Conservative 34; Mismatches 6; Indels 5163; Gaps 389;

QY 1 M-----S-----S-----DEV-----N----- 9
DB 297 MGRIRSVYPVAPQECNNMHLSTMKCNHCDEVSWQDPLKATCBHCGTENLVIEGPTT 356

QY 10 -----F-LV-- 12
DB 357 CGVLPNAVVMPCAPQDEIGPEHSVADYHNHNIETRLKGGTRCFGCGVFAVGC 416

QY 13 Y-R-Y-----L-Q--E--S-----G-FSH----- 23
DB 417 YNKRAYVWPASADIGSGHTIGDQVETLNEDLLEILSRVNIIVGDF-HLNEVAI 475

QY 24 -----SAP--T-----F-----G---I----- 30
DB 476 ILASFSASTSAFIDTILKSLDYKSFKTIVESCNGYKVTGKPKVGAWNIQQRSVLPLCG 535

QY 31 ---E---S-----H-----IS-OS-----N-- 38
DB 536 FPSQAAGVRSIFARTLDAANHSIPDLQRAAVTILDGISEQSLRLVDAMVYTSDLTNSV 595

QY 39 -I-----N--GALVP--P-----A-----A-----LIS- 51
DB 596 IIMAYVTGGLVQTSQWLSNLLGTTEKLRPIFEWIEAKLSAGVEFLKDAWEILKELITG 655

QY 52 ---I---IQ-----K-----G-----L-----Q----- 58
DB 656 VFDIVRGQIQVADSNIKDCVKCFIDVNNKALEMCIDQVITAGAKRLSLNLEGEVFIQSKG 715

QY 59 -Y-----V-----V-----EA-B--V-S--- 65
DB 716 LYRQCIRGKEQLQLMLPKAPKEVTFLEGDHDTVLTSEEVVLKNGELEALETPVDSFTN 775

QY 66 -----I-----N-----ED----- 69
DB 776 GAIVGTPVCVNGLMLEIKDKEQYCALSPGLLATNVFRLKGGAPIKGVTFGEDTWEVQ 835

QY 70 -----GT----- 71
DB 836 GYKNVRITFELDERVDKLVNEKCSVTVESGTEVTFACVBAVAVKTLQPVSDLLTNMG 895

QY 72 -----LFD--G-----R-----P-----I----- 78
DB 896 IDLDWSVATFVFLFDDAGEENFSRMYCSFYPPDEEEDAECEBEIDETCBHEYGTED 955

QY 79 -----E--S-----L-----S-----L----- 83
DB 956 DYQGLPLEFGASAEVTVRVEEEDWLDTTTEQSEIEPEPEPTPEPFVNOFTGYLKLTDN 1015

QY 84 --I---D-----A-----V-----M----- 88
DB 1016 VAIKCVDIVKEAQSANPMVIVNANHLKHGGVAGALNKAATNGAMQKESDDYIKLNGPL 1075

QY 89 -----P-----D----- 90
DB 1076 TVGGSCLLSGHNLAKCLHVGPNLWAGBDIQLLKAAYENFNSQDILLAPLLSAGIFGAK 1135

QY 91 -----V-VOT-R-Q-----QA--Y-----R----- 100
DB 1136 PLQSLQVCVQTVRTQVYIAVNDKALYEQVMVDYLDNLKPRVEAPKQEPPTEDSKTEK 1195

QY 101 -----D-----D-----KL--A-----O----- 105
DB 1196 SVQKPDVVKPKACIDEVTTTLEETKFLTNKLLLFADINGKLYHDSQNMRLGEDMSFL 1255

QY 106 -----O----- 106
DB 1256 EKDAFVMVGDVITSGDITCVVIPSCKAGGTTEMLSRALKKVPVDEVITTYPGGCGAGYTL 1315

QY 107 -----HA-----A----- 109
DB 1316 EBAKTALKKCSAFYVLPSEAPNAKEEILGTVSNWNLRELAHAETRLKMLPICMDVRAIM 1375

QY 110 A-----A----- 111
DB 1376 ATQRYKIKIKQEGIVDYGVRFFFTYSKEPVASITKLNLSNEPLVTPIGVYTHGFNL 1435

QY 112 --AA-----A-----A-----A----- 117
DB 1436 EEAARCSRSLKAPAVVSVSSPDVTTYNGVLTSSSTSEEHFVETVSLAGSYRDSVSGQ 1495

QY 118 -----TN 119
DB 1496 RTELGVFEFLKRGDKIVYHTLESVPFHLDEVLSDKLKSLSLREVTKIKVFTTVDNTN 1555

QY 120 -----QQ-G-S---A---K--N--G----- 127
DB 1556 LHTQLVDMSTYGGQPGPTLDGADVTKIKPHVNHGKTFVFLPSDDTLRSEAFYHTL 1615

QY 128 -E-----N-----T----- 130
DB 1616 DESFLGRYSALNHTKKWKPQVGGILTSIKWADNNCYLSSVLLALQOLEVKFNAPALQEA 1675

QY 131 -----AN-----GE-----E-----N--GA-----H--- 139
DB 1676 YRARAGDAANFCAILAYSNKTVGELGVRETMTLHLOHANLESARVLNVVCKHCGQK 1735

QY 140 -----T---I----- 141
DB 1736 TTTLTGVEAVMYGTLSDYDNLKTGVSIPCVGGRDATOYLVOQESSFVMSAPPAEYKLOQ 1795

QY 142 -----AN-----N-----H-----T-----D-----MM----- 149
DB 1796 GTFLCANETGYQCGHYTHITAKETLYRIDGAHLTKMSEYKGPVTDVFKETSYTTIK 1855

QY 150 -----EV-----DG-----D-----VE--I-----P-----SN----- 160
DB 1856 PVSYKLDGVTYTEIEPKLDGYVKDKNAYVTEQPIDLVPTQPLPNASFDNFKLCSNTKFA 1915

QY 161 -----K-A-----VV----- 164
DB 1916 DDLNQMGTGFKPASRELSVTFPPDLNGDVVAIDYRHSASFKKGAKLLHKPIVMHINQAT 1975

165 QY -----LR----- 166
1976 Db TKTTFKENTWCLRLCMSTKPDVTSNSFEVLAVEDTQGMNDLACESQQPTSEEVENPTIQ 2035
167 QY -----GHE-----S-----E----- 171
2036 Db KEVIEDVKTTVEVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNLSIALG 2095
172 QY -----VF----- 173
2096 Db LKTIATHGIAAINSVPWSKILAVYKPELGOAAITTSNCAKRLAQRVENNMPVFTLLFQ 2155
174 QY -----I-----C-----A-W----- 177
2156 Db LCTFTKSTNSRIRASLPTTIKNSKSVKAKCLDAGINYKSPKSKLFTIANWMLLLLSI 2215
178 QY -----N-P-----V-----S-D-L 183
2216 Db CLGSLICVTAAGVLLSNFGAPSYCNGVRELYLNSNVTMTDFCEGSPFPCSICLSGLDSL 2275
184 QY -----L-V-S-----G-S----- 189
2276 Db DSYPALETIOVTISSYKLDLTILGLAAEWLAYMLFTKFFYLGLSAIMQVFFGYFASHF 2335
190 QY -----D-ST-----A 193
2336 Db ISNSLWMFIIISIQMAPVSAMVRMYIFFASFYIYKSYVHIMDGGCTSSITCMCYKENRA 2395
194 QY -R-----I-----WN-LSN-ST--SG----- 205
2396 Db TRVECTITVNGMKRSFYVYANGRGCKTHWNCL-NCDTFCTGTFISDEVARDSLQ 2453
206 QY -----PT-Q--L-----V-----L-----RH----- 213
2454 Db FKRPINPTDOSSYIVDSVAVNGALHLYFDKAGOKTYERHPLSHFVNLDNLRANNTKGS 2513
214 QY -----C-I----- 215
2514 Db PINVIVFDGSKCDESASKASYYSQLMCOPILLDDQALVSDVGDSTEVSKMFDAYVD 2573
216 QY -----R----- 216
2574 Db TFSATFSPMEKLKALVATAHSELAKGVALDGLVSTFVSAARQGVVDTDVTDKDVIECLK 2633
217 QY -----E-G-----G-----O----- 220
2634 Db LSHHSDLEVTGDCNNFMPLYNKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVK 2693
221 QY ----- 221
2694 Db DYNLSLSQLRKQIRSAKKNNIPRLTCATTQVQVNVITTKISKGGKIYSTCFKMLKA 2753
222 QY -----V-P-S-----N-K--D-VT--S-----LD- 232
2754 Db TLLCVLAALVCYIYVMPVHTLSIHDGYTNEIIGYKAIQDGVTRDIISTDDCFANKHAGFDA 2813
233 QY -----N-S-----E-----GT----- 238
2814 Db WFSQSGSYKNDKSCPVAAIITREIGFIVPGLPGTVLRAINGDFLHFLPRVFSVAVGNC 2873
239 QY -----L-----LA-----T-----GS-----Y-- 245
2874 Db YTPSKLIEYDFATSACVLAECTIPKDMGKPPYCYDTNLLEGSISYSSELPDTRYVL 2933
246 QY -DG-----F-A-----R-I-----W----- 252
2934 Db MDGSIQFPNTYLEGSRVVVTTFDABYCRHGTCESEVIGCLSTSGRWLNNEHRYALSG 2993
253 QY -----TK-----D----- 255
2994 Db VFCGVDMNLIANIFTPLVOPVGLDVSASVAVAGGIIAILVTCAAYYFMKFRFRVFEYNH 3053
256 QY -----G-----N-----LA-----S--- 260

3054 Db VVAANALLFLMSFTILCLVPAYSFLPGVSVFYLYLTFYFTNDVSPFLAHLOWFAMESPIV 3113
261 QY -----T-----L-----G----- 263
3114 Db PFWITAIYVFCISLKHCHWFFNNYLRKRVNFNGVTFSTFEAAALCTFLANKEMYLKLRS 3173
264 QY -----Q-----H-K-----G-----P----- 268
3174 Db TLLPLTQYRYLALYNYKYFSGALDTSYREAAACCHLAKALNDFSNSGADVLVQPQTS 3233
269 QY -I-----F--AL--K-----W----- 274
3234 Db ITSAVLQSGFRKMAFPFGKVGCMQVTCGTTTLNGLWLDLDDTVYCPRHVICTAEDMLNPN 3293
275 QY -----N-----K----- 276
3294 Db YEDLLIRKSNHSLVQAGNVQLRVIGHSMQNCILLRLKVDTSNPKTKPKYFVRIPQGTFS 3353
277 QY -----KG--N-----F-I-L--S-----A 284
3354 Db VLACYNGSPGVYOCAMRPNHTIKGSLNGSCSGVGFNDYDCVSCYMHMELPTGVHA 3413
285 QY -----VDK-----TTI-----I--W-----D----- 294
3414 Db GTDLEGKFGYFPVDROTAQAAGTDTTITENVLAWLYAAVINGDRWFLNRFTTLNDFNLV 3473
295 QY -----H-----TG-----E----- 299
3474 Db AMKYNIEPLTDQHDVILGFLPSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTP 3533
300 QY -----AK-----Q-O-----F-PF-- 306
3534 Db FVVROCSGVTFOGKFKIVKGTTHWMLITFLTSLILVQSTOWSLFFPVYENAFPLFTL 3593
307 QY -----H-----S--A-----PA----- 311
3594 Db GIMAIACAMLLVKHGHAFCLFLPLSLATVAYFNVMYPASVMWMRIMTWLELADTSLG 3653
312 QY --L-D-----V-D--W-----Q----- 317
3654 Db YRLKDCMYASALVLLILMTARTVYDDAARRVMTLNVITLVYKYVYGNALDAQISMWAL 3713
318 QY -----SN-----NT----- 321
3714 Db VISVTSYSGVTTIMFLARIVFCVVEYYPFLFITGNTLQICIMLVYCFGLYCCCYFGL 3773
322 QY FASC-----ST-----D----- 328
3774 Db F--CLLENYFRULTGVYDVLVSTQEPYMNYSQGLLPKPSIDAFKLNILKLLGIGKPKIC 3831
329 QY -----M-----CI--H----- 332
3832 Db VATVQSKNDVKCTSVLLSVLQQLRVESSSKLWAOQVQLHNDILLAKDTTEAFKMWSL 3891
333 QY -----V-CK-----L-----G----- 337
3892 Db LSVLLSMQGVADINRLCEMELNDRATLOAISEFSSLPYAAAYATAQEAYEQAVANGDSE 3951
338 QY -----Q----- 338
3952 Db VVLKCLKSLNVAKSEFDRDAAMQKLEKMAQDMQMTOMYKQARSEDKRAKVTSAMQTMFL 4011
339 QY -----D-----R--P--I-----K-----TF----- 345
4012 Db TMLRLKNDLANNIINNARDGCVPLNIIFLTAAKLMVWVVDYGTYKNTCDGNTFTYASA 4071
346 QY -----O-----GH-----T-----NE--V----- 352
4072 Db LWEIQVVDADSKHVQLSEINMDSFNLAWPLIVTALRANSVAVKLQNNELSPVALROMSC 4131
353 QY -----NA-----I-----KW-----D-----P-- 359

```
Db 4132 AAGTTQTACTDDNALAYNNNSKGRFVLALLSDHQDLKWARPPKSDGTGTIYTELEPPCR 4191
QY 360 ---T---G-NL---LA---S-CS--- 368
Db 4192 FVTDTGPKGVKYLIFIKGLNNLRGVLGSLAATVRLQNGNATEVPAHSTVLSPCAFAV 4251
QY 369 D---D---M---T--- 372
Db 4252 DPAKAVDYLAGGQPTINCVMKLTHTGTGQAITVTPEANMQESFGGASCCLYCRCHI 4311
QY 373 ---LK---I---W---S---M--- 378
Db 4312 DHPNPKGCDLKGKYQIPTTCANDPVGFTLRNTVCTVCGMMKGYGSCDQLREPLMQSA 4371
QY 379 ---K---Q--- 380
Db 4372 DASTFLNRUCGSAARLTGCGTGTSTDDVYRAFDIYNEKVAGFAKFLKTNCCRFQEKDEE 4431
QY 381 ---D---N---C---VHDL---Q--- 388
Db 4432 GNLLDSFVVVKRHTMSNYQHEETIYNLVKDCPAVAHDFKFRVGDGMVPHISRQRLTKY 4491
QY 389 ---A-H---N---KEI--- 394
Db 4492 TMADLVYALRHFDGNCDTLKEILVTYNCDDDDYFNKKDWDYFVENPDILRVYANLGERV 4551
QY 395 ---Y--- 395
Db 4552 RQSLKTVQFCAMRDAGIVGVUTLQNDLGNWYDFGDEQVAPGCGVPIVDSYSLLM 4611
QY 396 ---T---IKW--- 399
Db 4612 PILTLRALAAESHMDADLAKPLIKWDLKYDFTBEERLCIFDRYFKYQDQYHPNCINCL 4671
QY 400 ---S---PT---GP--- 404
Db 4672 DDCILHCANFNLFSTVFPTSGPLVRKIFVDGVPFVSTGYHRELGVHQNQDNLH 4731
QY 405 ---G---TNN---P--- 409
Db 4732 SSRLSPKELLVYAADPAMHAASGNLLDKRTTCSVAALTNNAVQTVKPGNKNDFYDF 4791
QY 410 ---NA---NL-M---L--- 415
Db 4792 AVSGFKFGSSVELKHFPAQDGNAAISDYDYRYNLPMTCDIRQLLFVVEVDKYFDC 4851
QY 416 ---A---SASF---DS--- 422
Db 4852 YDGGCINANOIVNNLDKSGAGFPFNKGRKARLYDSMSYEDQDALFAYTKRNVIPITILM 4911
QY 423 ---TV---R---L---W--- 427
Db 4912 NLKVAISAKNRARTVAGVSTCSTMTNRQHQKLLKLSIAATRGATVVGTSKFGYGGWHNML 4971
QY 428 ---DV---DR--- 431
Db 4972 KTVYSDVEPHLMGWDYPRCDRAMPNMLRIMASLVLRKHNTCCNLSHRFYRLANECAQV 5031
QY 432 ---G---IC---I--- 435
Db 5032 LSEVMCGGSLYVKPGTSSGDATTAYANSVFNICQAVTANVALLSTDGKNIADKYVRN 5091
QY 436 ---H---TLT-KH--- 441
Db 5092 LQHRLVECLYRNRDVEHDFVEFYALRHRFMSMILSDDAVCYNSNYAAQGLVASIKNF 5151
QY 442 ---Q---E---P---VY---S-V-A--- 449
Db 5152 KAVLYYQNNVFMSEAKWTETDLTKGHEFCQSQHTMLVKQGDYVLPYDPDRRIIAGC 5211
QY 450 ---F-S---P---D---G--- 454
Db 5212 FVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNQEYADVFLHYLQYIRKLHDELTHMLDM 5271

RESULT 13
Q6XA09 PRELIMINARY; PRT; 7191 AA.
ID Q6XA09
AC Q6XA09;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Nonribosomal peptide synthase.
GN Name=NRPS1;
OS Alternaria brassicae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Pleosporaceae; Pleosporaceae; Alternaria.
OX NCBI_TaxID=29911;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14727058;
RA Guillemette T., Sellam A., Simoneau P.;
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria
brassicae and flanking genomic sequences.";
RL Curr. Genet. 45:214-224 (2004).
CC -I- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
DR EMBL; AY246697; AAP78735.1; -.
```

DR InterPro; IPR010071; AA adenyl_dom.
 DR InterPro; IPR009081; ACP like.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR000977; DNA ligase.
 DR InterPro; IPR006162; Pplantne S.
 DR InterPro; IPR006163; Pp bind.
 DR Pfam; PF00501; AMP-binding; 4.
 DR Pfam; PF00668; Condensation; 9.
 DR Pfam; PF00550; PP-binding; 6.
 DR PRINTS; PR00154; AMPBINDING.
 DR TIGRFAMs; TIGR01733; AA-adenyl-dom; 4.
 DR PROSITE; PS00075; ACP DOMAIN; 7.
 DR PROSITE; PS00455; ACP BINDING; 2.
 DR PROSITE; PS00697; DNA_LIGASE A1; UNKNOWN 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN 3.
 SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.4%; Score 2618.9; DB 2; Length 7191;
 Best Local Similarity 8.6%; Pred. No. 3e-41;
 Matches 451; Conservative 48; Mismatches 11; Indels 4751; Gaps 360;

QY 1 M-----S-----IS-----S-----D-----E----- 7
 DB 1521 MAFRLAGSGRAPAGHAETLRLWASVLKLDAGAVGAEDSFRLGGDSISAMKLVTAAR 1580
 QY 8 ---V-N---F---L---V---Y---R----- 14
 DB 1581 KGVVLNVASVFAQPKLAWAATAVMLPSDDAAKPEADTLPMELLPASSRQAIVALAASE 1640
 QY 15 -----Y---LQE----- 18
 DB 1641 CDVFPDCIEDMYPCSKLOEGLVMTNKDPGYVVQPIYRLPADMDISRFKEAMKVVAAE 1700
 QY 19 -----S---GF-----S---H---S-----A-FT----- 27
 DB 1701 ASLRTRIVYEEHGLVQVVRREDIKWQSLPDIQHINETTRRLPAKNGAPLATFTLVGENT 1760
 QY 28 ---F-----G-----IE---S---H-----IS----- 37
 DB 1761 DSMFFVWTAHVYDGSWTALFRKVEANYRSGVHGMPATVPYSRFVKYISSLDQRQSDA 1820
 QY 38 -----NI-----NGAL---V---P-----P---AA---L--- 49
 DB 1821 FWLTQLDNVTAQFPQLPSDPHRVEANGQLLHTVLTNRNFGLEVTVPSMIRAAWGILLAT 1880
 QY 50 -----I-SII-----OK----- 55
 DB 1881 YSGSDDDVIMGETNSGREASVPGIESIIGPTITTAAPVRLRLNRSILTVHDKETQKQSSLS 1940
 QY 56 -----GLO-----Y-VE---AE---VS---I----- 66
 DB 1941 LPYQFAGLQHIKSLSPETAIACDFQSLGIEAGDDFPAESSLWNWVSANTIGTDFSYAF 2000
 QY 67 -----N---E-----D-----GT----- 71
 DB 2001 VFENKYNATGVQVEALPDDRVTQTLAQRVQVQDFLTQLQNGTDNLTLSLDDLINPA 2060
 QY 72 -----L-----F-----D----- 74
 DB 2061 DRKTISSWNSKVPFIIPRCIHVSIAEDQTLRPTACAIDAWDTGFMYSYELDERSSALAH 2120
 QY 75 -----G-RP-----I----- 78
 DB 2121 QLIRLGVRPKQFVPLCFDKSGWTIVAIIVLKGAFAFVPLDFEAPVLRLEIVSDIDADL 2180
 QY 79 -----E---S-----L-SL----- 83
 DB 2181 LLCAPQRELQCSIPCSMTVVDROATETVAGRLPSLVSHSDSPAFYITSGTGKPKGA 2240
 QY 84 -----I-----DA---V---M-----PD----- 90
 DB 2241 VVHTHWVTSSTAPAGWKISTASRVLQFASYPDACLIIEVFTLMGGTVCVPDQGSRT 2300

QY 91 -----VW-----O----- 93
 DB 2301 NDLVGVINRPNVWNAALTFSVVRMIVPSRIPQLETFLVCEAMSQQDLVTWADKVNLCNG 2360
 QY 94 -----T-----R----- 95
 DB 2361 YGPTCEAAVATSNIMTPHMRPNNLGRAVTARGWIVSRNNHHTLAPVCAIGELLEGGAVG 2420
 QY 96 -----O---QA-----Y-R-D--- 101
 DB 2421 AGYLNPEKTAQVFVQCARWCVGLMGDDISAPVRIYKTGLVKYNEDGTMLYLGRKDLQT 2480
 QY 102 -----KL---AQ-----O-----H----- 107
 DB 2481 KVRGQBLELSEVEHKLDDHVMQVSALASVPTTGPCAKRLVAIVSLQKHGHDHDTDKQLR 2540
 QY 108 -----A---A-----A-----A-----A-----A 113
 DB 2541 LLPQENASLENIATIRDLGCLERLPAYMIPSLWIAVERFPLMPGKMDRRCAIQMLEQMDQA 2600
 QY 114 -----A---A-----A---A---T-----NQ----- 120
 DB 2601 TYRLISAMGTDDAEVNVNGSPIEKLRIFAIVLNLNSVADVRLNQSFHLGGDSIAAMQVS 2660
 QY 121 -----QG----- 122
 DB 2661 SQCRAGFPISVQDIIRSKSISMASAVDLSQSSQVTTTEAKDYDLDFDLSPIQKVPFVAV 2720
 QY 123 -----SA-----KN-G-----E----- 128
 DB 2721 GDKHYFNQTELPRLSRNIETEELRSALTVLAKTHPMLRARFSEKNEAGAKQRIEKDVSS 2780
 QY 129 -----NTAN----- 132
 DB 2781 SFLRHHVQAGNDANLRPIIDHSQATLDVAKGPTFAIELFDVDDTFSSAIALVAHLII 2840
 QY 133 -----G-----E----- 134
 DB 2841 DVVSWGILLEDLQGLQLOPAPQSPMPYHMLHEQSLQATQESARRVFPVGDIA PGDLDY 2900
 QY 135 -----E---NG-A---H-----T 140
 DB 2901 WAMEGRPNVNGDVVEEDHLSTRDTMLLGAQDALATETIDILIAALFESFRKVPDRST 2960
 QY 141 IA-----NN-----H-----TDM----- 148
 DB 2961 IATHNEGHGRETFNNRQDLSRTVGWFSTVTPHLPVPLNEATDMISTIRWVDRFRNRTPD 3020
 QY 149 -----ME---V-----DG-----DV-----E-I 157
 DB 3021 KGRPYFAYRNLTGEGQTRFASHWPAEAVFNVYVGRLOSQDNKDLFTALSVDVDSREVGEDV 3080
 QY 158 P-----S-----N-----K-----AV-----V 164
 DB 3081 PRLALFDITAAVSGQAIKLSFGWNRNRQKEIRAWVGKCRQTLVDAVEELLQARQERSV 3140
 QY 165 -----L---R-----G-----H-E----- 169
 DB 3141 GNFKYLLLYNGISRLSAILPAGINLNDVEDIYPASPMOQGLLHTQSRHPELYTYHTVSQ 3200
 QY 170 --S-----E-----VFI----- 174
 DB 3201 VQSADGNPIDPRLAQWQVVVHRHQALRTIFIDSLAKGSKDQIVLKEKGRVQIILADC 3260
 QY 175 -----C-A-----W-----N----- 178
 DB 3261 DDSQVANLLRHQSSIDCREALPBRMSICKTKGRVWFKLESHVINDGTSVSNLLADLA 3320
 QY 179 -----P-----V-----SD-----L---V---SG-----S 188
 DB 3321 RAYARKLTRADAGPLYSDYIGYMLSRSSDADLAWKAHLGIEFCFLFPVNLNDGIPSPES 3380

QY 189 G-D---STARI-----W-----N 197
Db 3381 GSDVELGSTSRVQDFCKONGVTLNVNQLTWTALTLHYVGTFDVSGFLIASGRDIPVTN 3440
QY 198 LSE-----N-----ST---S---G-----P-----206
Db 3441 INEAVGCFVMVVSRLSFSDETTIAQLLEALQSTGSTEALSHOGCSLADIQHALLQPSLFN 3500
QY 207 ---T-Q-----LVL-----R-----212
Db 3501 TATFQRRSLSSDPEDTALIYEDMEAEADAGEYIVTVNADVTQOSITVDFGYSKDRILPSQ 3560
QY 213 -----H-----C-----I---215
Db 3561 AQNAETFFKILDSIVVCSASELTIGKLDILTESSIHQIMENPQLPPPIRRCLHDVHID 3620
QY 216 ---R---EG---G---QD-----221
Db 3621 QALTRPRTTKAVEGWDGTFYQDFDKITNQLAVHLOSIGVTTTETFPILFEKSSYAIVSM 3680
QY 222 -----221
Db 3681 IAIMKAGGAYVLPDPKHPQTRRLBELIEDVGSVULCSRGVHTTASEVAKTAVIVDORSIR 3740
QY 222 ---VP---S-----K-----226
Db 3741 KLGVPISSKPRTCATPDNAAAYCLFTSGTTGKPKGTIIPHQAFCTSAAAFTRRMNINATSR 3800
QY 227 ---D---VTS-L-----D-----W-----233
Db 3801 TFQFASYTFDASCIEILSALTVGATVCVPTEDDRMNAAGAIRKLRVNMSSLTPSVLGTI 3860
QY 234 -----N-----S 235
Db 3861 EPERVPLKTLVSGGEALSGPIKKWSNSTCFINAYGPTCSVVAATAYKSTLDHKLIVS 3920
QY 236 E-GTLIATGS---Y-----D-----G---FAR-----I---251
Db 3921 EPGT-IGTSGCRLWIVHPRNHDKLMFVSGVSELVIEGPTVARGYLNDVKTAKAFINDP 3979
QY 252 -W-----T---K-----DG--NL---AST---L-GQ-----H---265
Db 3980 AWAKTIFSSNNTFEAARMYKTGDLVRYNTDGSVNYIGRKDTQIKLANGQRIELGEIBFHVG 4039
QY 266 -----265
Db 4040 KNPPERVSARELVAPSNRSSAKALAVFAVVOQDAIDGQSVQVQASTDLPAADDLLLLPL 4099
QY 266 ---K-----G---P-IF---ALKW-----274
Db 4100 SDELRDMCKNTENGLAGSLPSYMPAIFIPVTKLPWTSAGKLDNRRLSLVQNLBREWA 4159
QY 275 ---NKK-----G-----278
Db 4160 MYRLSTANKKPEITEAEKKIHKAVCSVLSPSSVGIDDSFVRLGGDSISSMRLVAMAH 4219
QY 279 -----N-----P-----I-----281
Db 4220 TEQMBLSFDIFKNPKLSLAKTGAQISKSQAQKVMQPPDLLPASLTRSDVISEVVQOC 4279
QY 282 ---LS---AG---V-----DKT-----289
Db 4280 QVSKEDLQDAYPTSSIQDALLTSLIKQAGAYVAQHVLAFLPKSLDMTKFKAWEESAIOEID 4339
QY 290 ---T-II-----W-----DA-----H-----T---297
Db 4340 ILRTRIITQPSGIFMOAVLRNDPVDWREAKSLKSAEDDASKIPPHLGHLAAVTLVTTS 4399
QY 298 GE-----A-----K-----Q-----302
Db 4400 GERYFWTLHLYDGSIVYLMQVQOQIYKSGVSTTPQTSYARFVEYLSSTSVSDSVY 4459
QY 303 ---QFP-----P-HS---A-----PA-----311

Db 4460 WRERLTGVNAYQPPRPSHATSSAPPNGQMFQHSKMIHAKNTDVTDPANAIARAALILAA 4519
QY 312 ---L-----DV-----314
Db 4520 YTGSDDDVFGETLAGRDVAMVTGITDVGPTLTTPVSRVKIDRGATVSDLLNTIATNITDR 4579
QY 315 -----D-W-----Q-----317
Db 4580 IPYQHGLSAILKALGEDMIAACDFQNLVLVQTEBEBELADSMWSVHDNEEQNFYPLVI 4639
QY 318 ---S-----N-----319
Db 4640 ECKMGLSKTEVLAHFADANVISLWHVORLIYQFETVLIQSATHVRHIAVLSQDQKQVR 4699
QY 320 --N-----T---F-----AS-----324
Db 4700 KMNAYEPRLIDDTVPSPFPKKAASQPTTAVTADFGEFSYGELSALASQAQBLVKFGAG 4759
QY 325 ---C-----S-TD-----328
Db 4760 PECLIPICVDKSRWAIVAILAILISGAGYVPLSPDTPASRHLHIVETCKASIVLCSPKYT 4819
QY 329 ---M-----C-----I-H---332
Db 4820 HRFVEMVGHVFSVSETAIRQLPTSSISLSQRAKNNICVIFTSGTGLPKGVVIEHKS 4879
QY 333 ---VCKLG---Q-----D-----R-----P-----341
Db 4880 VSSAAACE-GLHITPSTRVFOFCFLFDVSVGETLTVLLRGATICVPSDEQRTNLAAA 4938
QY 342 IKT---F-----QG-----HTN---350
Db 4939 V-TDNLNANWFLTPSVASTLEGKPSVPTLETLLVVGGEAMTSDVVVKWATGVNLH-NGYGP 4996
QY 351 -E-V---N---A---I---K---W---D---P---T---360
Db 4997 TEGTVAIGNDHVSAQRDPSNIGHPLKSGRAWLTNSDNPHELAPIGATABLECLEGPLLAR 5056
QY 361 G-N-----L-----L-----A---S-----366
Db 5057 GYLNDPKRTSEAFIEAPAFKPNFSSNSERIYRTGDLVQYAADGSIQYMRGRKNQIKLAG 5116
QY 367 -----C-----SD-369
Db 5117 QRIELDEIQAVHADNNHVQVVQLPKVGPCYKTLTVVVSPPGTAASAGSDWRILSDT 5176
QY 370 ---D---M-----TL-----K-----374
Db 5177 ESLSQINRRARDRLADLVPSYVMPFIWIAVPRIPTLASAKLDKKQVGLWLEGMDEALYQRI 5236
QY 375 -----IW-----S-MK-----379
Db 5237 MGAELPEDMEGFAAALTVLRGIAWAKVLRNPVEDVKPSKSWLSLGGDSISAMKLLAKCRS 5296
QY 380 -----Q-D-----381
Db 5297 EGINLNLQILRAKSLSHLAADVKSVVILDHEKQNDRFALSPIORFYVEAGSIENSTH 5356
QY 382 ---N-----C-----V-----384
Db 5357 FNOGSTLRILRVQPAWVQALNSIVECHSLMRARFSKDNNGOWQOLVMSKVSQSYAFTA 5416
QY 385 HD-----L---Q---AH-----390
Db 5417 HDVSTASAAGAIISSTQKSLDIRTGPFAVDLFLNKGHQLFMAAHLHLVLDVVSWGILLG 5476
QY 391 -----390
Db 5477 DLEDLGSFVTLPRSLPFTQWCKMQTSASBITOQLTVKNQPLVVEPANFAPFWMGDVR 5536
QY 391 -N-----KE-----IY-T-----I-K-----398

Db 5537 PNVYGDSEDFVIDKETSAMAFDNHHVYKTDLDVLDILLAAILHSESRVFINKRAPTLENE 5596
Qy 399 ---W---S---P---T---G-P-- 404
Db 5597 SHGREVWGSNLDLSRTVGWFTLYPVTPVPIDDDEDEVIHTRQVKDTRRKVASNGRPYF 5656
Qy 405 ---G---TN---N---P--- 409
Db 5657 AHRVLTEDGKERFANHAPMEVFLNYLGRQESGSHDLSLLSPTQVEGDDDETSYGVKTSR 5716
Qy 410 ---N---AN---L---M- 414
Db 5717 MALFEISASVTEGOQLSFMWNRYSKKQKGIIRWIAECQRTLEBIAIEIAELAKTRDPQMTA 5776
Qy 415 ---LAS---A---SPD-----ST- 423
Db 5777 DFPLPLESRLDRVLKTLPHAGVPSFQVEDMVPSPIDGMILSQIKSPESYWSSTT 5836
Qy 424 ---VR---L-W---D---V-DR---GI---C 434
Db 5837 FEVRSKRGPDASKVVDGKQVVARHPALRTIFIDSVCKGGVFDQIVVKSPDSGIVTYKC 5896
Qy 435 ---I-HT---L---T---K---H--- 441
Db 5897 ADAELATLLESIRHSSSLNGKKPVLPQAAVAVQTSGGKIFVKVIVNHAVIDGGSLGVIGQ 5956
Qy 442 ---QE---PVVS---VA---F---S-P-- 452
Db 5957 DLQAYEGRUSEDGLYSAYIKYLRALPAEDAIAYKAKLRGVSPCYFPTTPRDPSPRQ 6016
Qy 453 ---D---D---G-R---YLAS--- 459
Db 6017 LOSLDMRFTPELDHDAESSNVITIANILLAAMALVLRSTNSDDVCYGLTSGRNVPID 6076
Qy 460 ---GSF---D---K---C--- 465
Db 6077 RIENAVGAFINMLVSRTELKATSLSLEIVENVQSDFGVSGMPHQHCSLAQFOHDLGLSGKS 6136
Qy 466 ---V---H--- 467
Db 6137 LFNTAVSIQRSSPEELTSDSGIEFEQLDGHDPSEFAITVNIDATRDGEVRYTWSDA 6196
Qy 468 --- 467
Db 6197 VTDEAKNVSTLMAKILVQLSNPKQTIAELDIVVKGPAQPAVSRNLNPKPRPSILRSS 6256
Qy 468 ---I---W---N--- 470
Db 6257 SSISRSSTSPRTPRIITFPDLAPALPAETPDWSSLIRSVSEMPQIVEQIVAQNKLS 6316
Qy 471 ---TGAL---VHSYR--- 481
Db 6317 TEPTSATIDQMTQMLTKRTKISORGRPSIDDDAGSPRAASHRRAMSIASNAENRI 6376
Qy 482 ---G---T---GG---I 486
Db 6377 QTAADMVATLGLVLAETASKVADPVEKLLNLAELLENVEETIEQDDSFQGLGDSLI 6436
Qy 487 ---F---EV-C--- 490
Db 6437 AMRLVGAAREEGLSMTVADVFNKPTFADMARVVRVAGEVIDEVMSRAGGSVAGRSAGQS 6496
Qy 491 ---W---NA--- 493
Db 6497 RSLRHDRAPIWSEFQDIMSDPNADTKSIAPSEMPAEPGLNSRDSMTWFQWQGLTNNQAR 6556
Qy 494 ---A---GD---KV---G--- 499
Db 6557 PAASREVSQKSLATHITIQEGVESSVNRVSMLGDPNVESVISKVQVKGGISDVFPVPTDF 6616
Qy 500 ---A---S---A--- 502
Db 6617 QSLAITGLMESKMWLNLYFYLDGGDPLDLRLKLAQAYRMVHAFDILRTVFPVYGDRLQV 6676

Qy 503 ---SD---G----- 505
Db 6677 VLRLQLOPEFYHQTDDDDITFTKDLRKQKRENGPKLGEAFIQFVAKQKQOTGRYRIFMRL 6736
Qy 506 S-----VC---VL---D---L-----R---K 514
Db 6737 SHAQYDGVCMKILNALQDGYNGLPVSSAPSGNFVRETAK 6777

RESULT 14

AAP78735
ID AAP78735 PRELIMINARY; PRT; 7191 AA.
AC AAP78735;
DT 01-APR-2004 (T-EMBLrel. 27, Created)
DT 01-APR-2004 (T-EMBLrel. 27, Last sequence update)
DT 01-APR-2004 (T-EMBLrel. 27, Last annotation update)
DE Nonribosomal peptide synthase.
GN NRPS1.
OS Alternaria brassicae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.
OX NCBI_TaxID=29911;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14727058;
RA Guillemette T., Sellam A., Simoneau P.;
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria
brassicacae and flanking genomic sequences.";
RL Curr. Genet. 45:214-224(2004).
DR EMBL: AY246697; AAP78735.1; -;
SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.4%; Score 2618.9; DB 2; Length 7191;
Best Local Similarity 8.6%; Pred. No. 3e-41;
Matches 451; Conservative 48; Mismatches 11; Indels 4751; Gaps 360;

Qy 1 M-----S-----IS-----S-----D-----E----- 7
Db 1521 MAPFLAGKSGRAPAGHAETVLARLWASVLKLDAGAVCAEDSPFRFGDGSISAMKLVTAAR 1580
Qy 8 ---V-N---F---L---V-----Y---R--- 14
Db 1581 KDGVLNVASVFAQPKLAWATAVMLPSDDAAKPEADTLPMELLPASSRQAIVALAAE 1640
Qy 15 ---Y---LOE----- 18
Db 1641 CDVFPDCIEDMYPCKLOEGLVMTNKPDTYVQPIYRLPADMDISRPEAKMVKVVAE 1700
Qy 19 ---S---GF-----S---H---S-----A-PT----- 27
Db 1701 ASLRTRIVYSEHGFLQVVVRREDIKWQSLPDQIHNETTRRLPAKNGAPLATFTLVGENT 1760
Qy 28 ---F-----G-----IE---S-H-----IS-----QS-- 37
Db 1761 DSMFFVWTAHAYDGCWSLTALFRKYANVRSGVHSPATVPYSRPVKYISSLDQSQDA 1820
Qy 38 ---NI-----NGAL---V---P---P---AA---L--- 49
Db 1821 FWLTQLDNVTAQFPOLPSPDHRVEANGQLHTVLLTRNPGLEVTVPMSIRAAWGILLAT 1880
Qy 50 ---I-SII-----OK----- 55
Db 1881 YGSSDDVINGETNSGREASVPGIESIIGTITITAPVRLNRLSLTVHDYVKETQKQSSLS 1940
Qy 56 ---GLQ-----Y-VE-----AE-----VS---I--- 66
Db 1941 LPYQFAGLQHISKLSPEATAICDFQSFGLGIEAGDDFDASSLWNVMSANTIGTDFPSYAF 2000
Qy 67 ---N---E-----D-----GT----- 71
Db 2001 VFNCNKVATGVOVEALFDDRVRVQVTLAQRMVQOQDFILTQLNGTNIILRSLODLDLINPA 2060

QY 72 -----L-----F-----D----- 74
Db 2061 DRKTISSMNSKVPPIIPRCHSVIAEDQALRPTACAIADWDGFMYSRELDERSALAH 2120
QY 75 -----G-RL-----I----- 78
Db 2121 QILRLGVRPKQFVPLCFDKSGMTIVAIIVLAKAGAFVPLDFEAPVRLREIVSDIDADL 2180
QY 79 -----E-----S-----L-SL----- 83
Db 2181 LLCAPQYRELQCSIPCTSTWVDRQATETVAGRLPSLPSVHSDSPAYFYTSGTGPKGA 2240
QY 84 -----I-----DA-----V-----M-----PD----- 90
Db 2241 VVHHTHWTSSTAFAGWKISTASRVLPASVYTFDACLIEVFSTLMQGGVVCVPDQGSRT 2300
QY 91 -----VV-----Q----- 93
Db 2301 NDLVGVINRPNVNAALTPSVVRMIVPSEIPQLETFLVGEAMSQDQDLVTWADKVNLGNG 2360
QY 94 -----T-----R----- 95
Db 2361 YGTECAAVATSNIMTPHMRPNHNLGRAVTARGWIVSRNHHHTLAPVGAIGELLLEGAVG 2420
QY 96 -----Q-----QA-----Y-R-D----- 101
Db 2421 AGYLNNPETAQVVFQARWCVCLMGDDISAPVRIYKTGDLVKYNEDGTMLYLGRKDLQT 2480
QY 102 -----KL-----AQ-----Q-----H----- 107
Db 2481 KVRGORLELSEVHEKLLDDHVMQSALASVPTTGPCAKRLVAIVSLQKHGDHDTDTDKQLR 2540
QY 108 -----A-----A-----A-----A-----A-----A 113
Db 2541 LLPQENASLIATIRGDLERLPAYMIPSLWIAVERPPLMPSGKMDRRCAIOWLEOMDQA 2600
QY 114 -----A-----A-----A-----T-----NQ----- 120
Db 2601 TYRLISAMGTDDAEVNVNGSPIEKRLRAIFAKVLNLSVADVRLNQSFHLGGDSIAAMQVS 2660
QY 121 -----QG----- 122
Db 2661 SOCRAGQFPISVDIIIRKSISAMASAVDLSSQSVTTEAKDYDLPDLSPQKVFDDAV 2720
QY 123 -----SA-----KN-G-----E----- 128
Db 2721 GDKHNYFNQTELPRLSRNIETEELRSALTVLAKTHPMLRARFNSKNEAGAWKQRIEKDVS 2780
QY 129 -----NTAN----- 132
Db 2781 SFLRLHHVQAGNDANLRIIDHSQATLDVAKGPTFAIELFDVDDTFSQAIALVAHLLII 2840
QY 133 -----G-----E----- 134
Db 2841 DVVSGILLLEDLQGLQLOPQSPMPYHMLWELQSLQTESARRVFPVGDIAPDGLDY 2900
QY 135 -----E-----NG-A-----H-----T 140
Db 2901 WAMEGRPNVNGDVEEDLHLSTRDTMLLGAQDALATETIDILIAALFSSFRKVPFDRST 2960
QY 141 IA-----NN-----H-----TDM----- 148
Db 2961 IAIHNEGHGRETTNNRQDLRSRTVGWSTVTPHLPVPLNEATDMISTIRWDRFRNRTPD 3020
QY 149 -----ME-----V-----DG-----DV-----E-I 157
Db 3021 KGRPYFAYRNLTTEGOTRFASHWPAEAVFNVYGRLOSQDNKQGLFTALSDVDSREVGEDV 3080
QY 158 P-----S-----N-----K-----AV-----V 164
Db 3081 PRALAFDITAASQGAIKLSFGWNRNMRQKEIRAWVGKRCQTLVDVAVEELLQARQERSV 3140
QY 165 -----L-----R-----G-----H-E----- 169

Db 3141 GNFKYLPLLYNGISRLSAILPAGINLNDVEDIYPASPMOQGLLHTQSRHPELYTYHTVSQ 3200
QY 170 -----S-----E-----VFI----- 174
Db 3201 VOSADGNPIDPRRLAEAWQVVRHQAALRTIIFIDSLAKDGSKDQIIVLKEKPRVQIADLC 3260
QY 175 -----C-A-----W-----N----- 178
Db 3261 DDSQVANLLRHQSSIDCREALPPHRMSICKTKTGRVWFVKLELSHVINDGTSVSNLLADLA 3320
QY 179 -----P-----V-----SD-----L-----L-----V-SG-----S 188
Db 3321 RAYARKLTRADAGPLYSDYIGYMLSRSSDADLAYKHAHLSGIEPCLPVPLNDGIPSPES 3380
QY 189 G-D-----STARI-----W-----N----- 197
Db 3381 GSVDELSTSRVQDFCKONGVTLSNVQLTWTALTYHYVGTDFVSFGLIASGRDIPVTN 3440
QY 198 LSE-----N-----ST-----S-G-----P----- 206
Db 3441 INEAVGCFVNVVSRLSFSDETTIAQLLEALQGTSTEALSHQGSCLADIQHALQLSLFN 3500
QY 207 -----T-Q-----LVL-----R----- 212
Db 3501 TAFTFORSLSDPEDTALIYEDMEADAGEYIVTVNADVTQGSITVDFGYSKDRILPSQ 3560
QY 213 -----H-----C-----I----- 215
Db 3561 AQNMATFKIILDSIVVCSASELTIGKDLITESSIHQIMENWNPQLPPPIRRCUHDVHID 3620
QY 216 -----R-----EG-G-----QD----- 221
Db 3621 QALTRPRITTKAVEGWDGTFTYQDFDKITNQLAVHLQSIGVTTETFPVILPEKSSYAIVSM 3680
QY 222 -----N-----K----- 221
Db 3681 IAIMKAGGAYVPLDPKHPQTRLRELIEDVGASVVLCSRGVHTTIASEVAKTAVIVDORSIR 3740
QY 222 -----VP-S-----N-----K----- 226
Db 3741 KLGVPISKPRTCATPDNAAAYCLFTSGTGKPKGTII PHQAFCTSAATFRRMNINATSR 3800
QY 227 -----D-----VTS-L-----D-----W----- 233
Db 3801 TFQFASYTFDASCIELLSALTVCATVCVPTEDDRMNAAGAIRKLVNMSLLTPSVLGTI 3860
QY 234 -----N----- 235
Db 3861 EPERVPLKTLVSGGEALSGPILKKWSNSTCFINAYGPTCSVVAATAVYKSTLDHKLIVS 3920
QY 236 E-GTLLATGS-----Y-----D-----G-FAR-----I----- 251
Db 3921 EPGT-IGTSGCRLWIVHPNRHDKLMPVSGVSELVIEGPTVARGYLNDEVKAKAFINDP 3979
QY 252 -W-----T-----K-----DG-NL-----AST-----L-GQ-----H----- 265
Db 3980 AWAKTIFSSNNTFEAARMYKTGDLVRYNTDGSVNYIGRKDTQIKLNGQRIELGEIEFHVG 4039
QY 266 ----- 265
Db 4040 KNFPERVQSAVELVAPNSRSSAKALAVFFAVVQDQAI DGEQSVQASTDLPAADLPL 4099
QY 266 -----K-----G-----P-IF-----ALKW----- 274
Db 4100 SDELDMCKNTENGLAGSLPSYMPIAIFTVTKLPWTSAGKLDNRNLSRVQNLSETMA 4159
QY 275 -----NKK-----G----- 278
Db 4160 MYRLTSIANKKPIEAEKKHAKVCSVLSPSSVIGIDDSFVRLGGDSISSMRLVAMAH 4219
QY 279 -----N-----F-----I----- 281

Db 4220 TEOMELSFIDIFKNPKLSLAKIAQISKSQAEKVMQPFDPDLPASLTRSDVISEVVQOC 4279
QY 282 -----LS---AG---V---DKT----- 289
Db 4280 QVSKEDQDAYPTSSLODALLTUSIKOGAGVAVQHVLPALPKSLDMTKFKAANESAIQEID 4339
QY 290 ---T-II---W---DA---H---T--- 297
Db 4340 ILTRIIQMPGIFMQAVLRNDPVDREAKSLKSAEDDASKIPPHLGCHLAAVTLVTTPS 4399
QY 298 GE-----A-----K-----Q----- 302
Db 4400 GERYFVWTLHLYDGSIVLMQORVQOYYSKGVSTTPQTSYARFVBYLSSTSVDSDVY 4459
QY 303 ---ORP---F-HS---A---PA--- 311
Db 4460 WRERLTGNVAYQPPRPSHATSSAPPNGOMFOHSMKIAHRKNTDVTAPANARAAWALILAA 4519
QY 312 ---L---DV--- 314
Db 4520 YTGSDVVFGETLAGRDVAMGITDVCPTLTTPVPSVKIDRGATVSDLLNTIATNITDR 4579
QY 315 -----D-W-----Q----- 317
Db 4580 IPYQHGLSAIKALGEDMIAACDFONLLVIQTENEELADSMWSVHDNEEQGNFTYPLVI 4639
QY 318 ---S---N--- 319
Db 4640 ECKMGLSKTEVLAHFANDVLSLHVORLIYQFETVLQLOSATHVRHIAVLSQDKQLVR 4699
QY 320 ---N---T---F---AS--- 324
Db 4700 KWAYEPRLLDDTVPSLFFPKKAASQPTTAVTAFDGEFSYGELSALASQLAQELVKFAG 4759
QY 325 ---C---S-TD----- 328
Db 4760 PEGLPICVDKRWAIIVAILAILISGAGYVPLSPTDPASRHLHIVETCKASIVLCSPKYT 4819
QY 329 ---W---C-----I-H--- 332
Db 4820 HRFVEMGVHVSSETAIRQLPTSSLSORAKNNICYIFTSGSTGLPKGVVIEHKS 4879
QY 333 ---VCKLG---Q---D---R---P--- 341
Db 4880 VSSSAALCE-GLHITPTSRVFCFSFLFDVSVGETLTVLRGATICVPSDEQRTNLAA 4938
QY 342 IKT-----F-----OG-----HTN----- 350
Db 4939 V-TDLNANWAFLPVSAVSTLEGPKSVPTLETLVVGGEAMTSDVVDKWATGWNLH-NGYGP 4996
QY 351 -E-V---N---A---I---K---W---D---P---T--- 360
Db 4997 TEGTVFAIGNDHVSAQRDPSNIGHPLKSGRAWLTNSDNPHELAPIGATABELCLEGPILLAR 5056
QY 361 G-N---L---L---A-S----- 366
Db 5057 GYLNDPKRTSEAFIAPAFKNFSSNESRIYRTGDLVQYAADGSIQYMGKKNQIKLAG 5116
QY 367 -----C-----SD----- 369
Db 5117 QRIELBEIQAVVHADNNHVQVQLPKVGPCTKKLTVVVSFPGTAASNAGSDWRRILSDT 5176
QY 370 -----D---M---TL---K----- 374
Db 5177 ESLSQINRARDRLADLPVSYNVPFIWIAVPRIPTLASAKLDKQVGLWLEGMDREALYQRI 5236
QY 375 -----IW-----S-MK----- 379
Db 5237 MGAELPEDEMGPGNAALTVLRGIVAKVLPVPVEDVKPSKWSLGGDSISAMKLLAKCRS 5296
QY 380 -----Q-D----- 381
Db 5297 EGINLNLINILRAKSLHLAADVKSVVILDHEQNDRPPALSPIQRFYVEAGSIENSTH 5356

QY 382 -----N---C-----V----- 384
Db 5357 FNOSSTLIRLYVQPAWQBALNSIVECHSMLRPARFSKNNQWQQLVMSKVGSGSYAFTA 5416
QY 385 HD-----L---Q---AH----- 390
Db 5417 HDVSTASAAGAILSSQKSILDIRTPGFVAVDLFNLAGHQLFMAAAHHLVIDVWSWGILLG 5476
QY 391 ----- 390
Db 5477 DLEDLLSGPVTLPRLSPFQTMCKMQTSASBITQOLTVKNQPLVVEPANFANFAGWMDVR 5536
QY 391 -N---KE---IV-T---I-K----- 398
Db 5537 PNVGSESRDEFVIDKETSAMAFDNHHVTKTDLVDILLAILHSFRVFINRRKAPTUFNE 5596
QY 399 ---W---S---P-----T---G-P--- 404
Db 5597 SHGREVWDGNSLDLSRTVGWFTTLYPVTVPIDDDDEVIHTLRQKDXTRRKVASNGRPYF 5656
QY 405 ---G---TN---N-----P----- 409
Db 5657 AHRYLTEDGKERFANHAPMEVLFNYLGRQEQSGHSDLSLPTQVEGDDDETSDVGKTSR 5716
QY 410 -----N---AN-----L-----M- 414
Db 5717 MALFEISASVTEGQIQLSFMWNRYSKNQKIRRIWAEACORTLBEIEIAELAKTORPQPTMA 5776
QY 415 ---LAS---A---SFD-----ST- 423
Db 5777 DPPLLPLESVRLDRVLKTLPHAGVPSFQVEDMYPCSPIDQGMILSQIKSPESYSSTT 5836
QY 424 -VR---L-W---D---V-DR-----GI---C 434
Db 5837 FEVRKRGVPDASKVVDGKQVVARHPALRTIPDSVCRGGVFDQIVWKSPPDSGIVTYKC 5896
QY 435 ---I-HT---L---T---K---H----- 441
Db 5897 ADELATLLESINHSNGKKPVLPHQAAVOTSSGKIFVKVIVNHAVIDGSSLGVIGQ 5956
QY 442 ---QE---PVYS---VA-----F---S-P--- 452
Db 5957 DLQEAVEGRLEDGPLYSAVIKYLRALPAEDAAYWAKLGRVSPCVFPTTPRDPSPKPRQ 6016
QY 453 ---D---G-R-----YLAS----- 459
Db 6017 LQSLDMRFRFDELHDLAESSNVTIANILLAAWALVLRSYTNSSDVCYGYLTSGRNVPID 6076
QY 460 ---GSF---D---K---C----- 465
Db 6077 RIENAVGAFINMLVSRIELKATSLSLLEIVENVOSDFVGSMPHOCHSLAQPHDLGLSGKS 6136
QY 466 ---V---H----- 467
Db 6137 LFNATVSIQRRSSPEELTSDSGIEFEQDGDHPSEFAITVNIDATRDDSGVFTYWSDA 6196
QY 468 ----- 467
Db 6197 VTDGEAKVSTLMAKILVQLSNPKQTIABLDIVVKRPAQPAVSRNLNPKRPSILRSS 6256
QY 468 -----I---W-----N--- 470
Db 6257 SSISRSRSTPPRTPRITFPDLAPAAPLPAETPDWSSLIRSIVSEMVQPIVEQIVAKNLS 6316
QY 471 -----TQ-TGAL-----VHSYR----- 481
Db 6317 TEPTSATIDMTQTMGLTRKTSISQGRPSIDTAGSPRAASIHSSRRASIASNAENRI 6376
QY 482 ---G-----T-----GG---I 486
Db 6377 QTAADMVATLGLVATEASSKVAPDFVEKLLNLWAELEWVEETIEQDDSPFOLGDSII 6436

QY 487 -----F-----EV-C----- 490
Db 6437 AMRLVGAAREGLSMTVADVFNKPTPADMARVVRVGEVIDEVMRSAGGSVAGRSAGQS 6496
QY 491 -----W-----NA----- 493
Db 6497 RSLRHDRAPIWSEPDIMSDPNADTKSIAPSEMPAEPGLNSRSDTMFORWOGLTTNOAR 6556
QY 494 -----A-----GD-----KV-----G----- 499
Db 6557 PAASREVQSLSLATHIQEGVESVNRVSMLGDPNVSIVSKQVFKGSDVFPVTFD 6616
QY 500 -----S-----A----- 502
Db 6617 QSLAITGTLMSKWLNLVYLDGDPDLRLKLRQAAYRMVHAFDILRTVPVPGDRFLQV 6676
QY 503 -----SD-----G----- 505
Db 6677 VLKLOPEFYHGTDDIDFTTKDLRKQRENGPKLGEAFIQFVVAQKQGRYIFMRL 6736
QY 506 S-----VC-----D-----L-----R---K 514
Db 6737 SHAQYDGVCMKSLNALQDNGLVPVSSAPSGFNFRYAK 6777

RESULT 15
LR3B MOUSE
ID_LR3B_MOUSE STANDARD; PRT; 4599 AA.
AC Q9J1I8; Q8B2D3; Q8B2M7;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Low-density lipoprotein receptor-related protein 1B precursor (Low-density lipoprotein receptor-related protein-deleted in tumor) (LRP-DIT).
DE Names=Lrp1b; Synonyms=lrp1dit;
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RA Yaklichkin S., Lisitsyn N.;
RT "The structure of the mouse homologue of the human candidate tumor suppressor gene Lrp1b/LRP-DIT";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 3770-4599 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12456851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Oosato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bratt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King P.B.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,

RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RT Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002)."
CC -!- FUNCTION: Potential cell surface proteins that bind and
CC internalize ligands in the process of receptor-mediated
CC endocytosis.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the LDLR family.
CC -!- SIMILARITY: Contains 17 EGF-like domains.
CC -!- SIMILARITY: Contains 32 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 25 LDL-receptor class B domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF270884; AAF81758.1; -;
DR EMBL; AK035795; BAC29188.1; -;
DR EMBL; AK034122; BAC28594.1; -;
DR HSSP; Q07954; 1CR8.
DR MGD; MGI:2151136; Lrp1b.
DR InterPro; IPR011044; Amine_DH_B_like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR Pfam; PF00008; EGF; 15.
DR Pfam; PF00057; Ldl_recept_a; 32.
DR Pfam; PF00058; Ldl_recept_b; 33.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00179; EGF_CA; 4.
DR SMART; SM00192; LDLA_32.
DR SMART; SM00135; LY; 36.
DR PROSITE; PS00010; ASX_HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS50026; EGF_3; 9.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS01209; LDLRA_1; 27.
DR PROSITE; PS50068; LDLRA_2; 32.
KW Calcium-binding; EGF-like domain; Endocytosis; Glycoprotein; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 4599 Low-density lipoprotein receptor-related
FT protein 1b.
FT DOMAIN 21 4444 Extracellular (Potential).
FT TRANSMEM 4445 4467 Potential.
FT DOMAIN 4468 4599 Cytoplasmic (Potential).
FT DOMAIN 31 70 LDL-receptor class A 1.
FT DOMAIN 76 114 LDL-receptor class A 2.
FT DOMAIN 120 154 EGF-like 1.
FT DOMAIN 155 194 EGF-like 2, calcium-binding (Potential).
FT DOMAIN 195 336 LDL-receptor class B 1.
FT DOMAIN 255 336 LDL-receptor class B 2.
FT DOMAIN 338 380 LDL-receptor class B 3.
FT DOMAIN 382 424 LDL-receptor class B 4.
FT DOMAIN 568 609 LDL-receptor class B 5.
FT DOMAIN 611 655 LDL-receptor class B 6.
FT DOMAIN 657 705 LDL-receptor class B 7.
FT DOMAIN 707 754 LDL-receptor class B 7.
FT DOMAIN 794 834 EGF-like 3.
FT DOMAIN 844 882 LDL-receptor class A 3.
FT DOMAIN 885 923 LDL-receptor class A 4.
FT DOMAIN 926 963 LDL-receptor class A 5.
FT DOMAIN 966 1003 LDL-receptor class A 6.
FT DOMAIN 1005 1043 LDL-receptor class A 7.

DOMAIN	1052	1089	LDL-receptor class A 8.	FT	DISULFID	180	193	By similarity.
FT	1094	1132	LDL-receptor class A 9.	FT	DISULFID	798	809	By similarity.
FT	1135	1174	LDL-receptor class A 10.	FT	DISULFID	805	818	By similarity.
FT	1172	1213	EGF-like 4.	FT	DISULFID	820	833	By similarity.
FT	1300	1345	LDL-receptor class B 8.	FT	DISULFID	845	857	By similarity.
FT	1347	1388	LDL-receptor class B 9.	FT	DISULFID	852	870	By similarity.
FT	1390	1435	LDL-receptor class B 10.	FT	DISULFID	864	881	By similarity.
FT	1437	1480	LDL-receptor class B 11.	FT	DISULFID	886	898	By similarity.
FT	1481	1521	LDL-receptor class B 12.	FT	DISULFID			
FT	1527	1570	EGF-like 5.					
FT	1618	1659	LDL-receptor class B 13.					
FT	1661	1703	LDL-receptor class B 14.					
FT	1705	1743	LDL-receptor class B 15.					
FT	1748	1788	LDL-receptor class B 16.					
FT	1834	1875	EGF-like 6.					
FT	1922	1963	LDL-receptor class B 17.					
FT	1965	2006	LDL-receptor class B 18.					
FT	2008	2050	LDL-receptor class B 19.					
FT	2052	2094	LDL-receptor class B 20.					
FT	2143	2183	EGF-like 7.					
FT	2239	2278	LDL-receptor class B 21.					
FT	2281	2328	LDL-receptor class B 22.					
FT	2330	2373	LDL-receptor class B 23.					
FT	2375	2416	LDL-receptor class B 24.					
FT	2417	2458	LDL-receptor class B 25.					
FT	2509	2548	LDL-receptor class A 11.					
FT	2551	2587	LDL-receptor class A 12.					
FT	2590	2626	LDL-receptor class A 13.					
FT	2629	2675	LDL-receptor class A 14.					
FT	2681	2717	LDL-receptor class A 15.					
FT	2719	2757	LDL-receptor class A 16.					
FT	2760	2800	LDL-receptor class A 17.					
FT	2804	2841	LDL-receptor class A 18.					
FT	2844	2885	LDL-receptor class A 19.					
FT	2890	2926	LDL-receptor class A 20.					
FT	2926	2967	EGF-like 8.					
FT	2968	3008	EGF-like 9.					
FT	3055	3097	LDL-receptor class B 26.					
FT	3099	3140	LDL-receptor class B 27.					
FT	3142	3184	LDL-receptor class B 28.					
FT	3186	3223	LDL-receptor class B 29.					
FT	3225	3266	LDL-receptor class B 30.					
FT	3274	3314	EGF-like 10.					
FT	3316	3353	LDL-receptor class A 21.					
FT	3356	3392	LDL-receptor class A 22.					
FT	3395	3432	LDL-receptor class A 23.					
FT	3435	3472	LDL-receptor class A 24.					
FT	3475	3511	LDL-receptor class A 25.					
FT	3514	3550	LDL-receptor class A 26.					
FT	3552	3588	LDL-receptor class A 27.					
FT	3593	3629	LDL-receptor class A 28.					
FT	3631	3668	LDL-receptor class A 29.					
FT	3673	3711	LDL-receptor class A 30.					
FT	3714	3752	LDL-receptor class A 31.					
FT	3761	3797	LDL-receptor class A 32.					
FT	3801	3843	EGF-like 11.					
FT	3844	3881	EGF-like 12.					
FT	3981	4037	LDL-receptor class B 31.					
FT	4039	4081	LDL-receptor class B 32.					
FT	4083	4126	LDL-receptor class B 33.					
FT	4213	4249	EGF-like 13.					
FT	4249	4290	EGF-like 14.					
FT	4286	4326	EGF-like 15.					
FT	4321	4363	EGF-like 16.					
FT	4392	4430	EGF-like 17.					
FT	32	45	By similarity.					
FT	39	58	By similarity.					
FT	52	69	By similarity.					
FT	77	90	By similarity.					
FT	84	103	By similarity.					
FT	97	113	By similarity.					
FT	159	169	By similarity.					
FT	165	178	By similarity.					
DOMAIN	1052	1089	LDL-receptor class A 8.	FT	DISULFID	180	193	By similarity.
FT	1094	1132	LDL-receptor class A 9.	FT	DISULFID	798	809	By similarity.
FT	1135	1174	LDL-receptor class A 10.	FT	DISULFID	805	818	By similarity.
FT	1172	1213	EGF-like 4.	FT	DISULFID	820	833	By similarity.
FT	1300	1345	LDL-receptor class B 8.	FT	DISULFID	845	857	By similarity.
FT	1347	1388	LDL-receptor class B 9.	FT	DISULFID	852	870	By similarity.
FT	1390	1435	LDL-receptor class B 10.	FT	DISULFID	864	881	By similarity.
FT	1437	1480	LDL-receptor class B 11.	FT	DISULFID	886	898	By similarity.
FT	1481	1521	LDL-receptor class B 12.	FT	DISULFID			
FT	1527	1570	EGF-like 5.					
FT	1618	1659	LDL-receptor class B 13.					
FT	1661	1703	LDL-receptor class B 14.					
FT	1705	1743	LDL-receptor class B 15.					
FT	1748	1788	LDL-receptor class B 16.					
FT	1834	1875	EGF-like 6.					
FT	1922	1963	LDL-receptor class B 17.					
FT	1965	2006	LDL-receptor class B 18.					
FT	2008	2050	LDL-receptor class B 19.					
FT	2052	2094	LDL-receptor class B 20.					
FT	2143	2183	EGF-like 7.					
FT	2239	2278	LDL-receptor class B 21.					
FT	2281	2328	LDL-receptor class B 22.					
FT	2330	2373	LDL-receptor class B 23.					
FT	2375	2416	LDL-receptor class B 24.					
FT	2417	2458	LDL-receptor class B 25.					
FT	2509	2548	LDL-receptor class A 11.					
FT	2551	2587	LDL-receptor class A 12.					
FT	2590	2626	LDL-receptor class A 13.					
FT	2629	2675	LDL-receptor class A 14.					
FT	2681	2717	LDL-receptor class A 15.					
FT	2719	2757	LDL-receptor class A 16.					
FT	2760	2800	LDL-receptor class A 17.					
FT	2804	2841	LDL-receptor class A 18.					
FT	2844	2885	LDL-receptor class A 19.					
FT	2890	2926	LDL-receptor class A 20.					
FT	2926	2967	EGF-like 8.					
FT	2968	3008	EGF-like 9.					
FT	3055	3097	LDL-receptor class B 26.					
FT	3099	3140	LDL-receptor class B 27.					
FT	3142	3184	LDL-receptor class B 28.					
FT	3186	3223	LDL-receptor class B 29.					
FT	3225	3266	LDL-receptor class B 30.					
FT	3274	3314	EGF-like 10.					
FT	3316	3353	LDL-receptor class A 21.					
FT	3356	3392	LDL-receptor class A 22.					
FT	3395	3432	LDL-receptor class A 23.					
FT	3435	3472	LDL-receptor class A 24.					
FT	3475	3511	LDL-receptor class A 25.					
FT	3514	3550	LDL-receptor class A 26.					
FT	3552	3588	LDL-receptor class A 27.					
FT	3593	3629	LDL-receptor class A 28.					
FT	3631	3668	LDL-receptor class A 29.					
FT	3673	3711	LDL-receptor class A 30.					
FT	3714	3752	LDL-receptor class A 31.					
FT	3761	3797	LDL-receptor class A 32.					
FT	3801	3843	EGF-like 11.					
FT	3844	3881	EGF-like 12.					
FT	3981	4037	LDL-receptor class B 31.					
FT	4039	4081	LDL-receptor class B 32.					
FT	4083	4126	LDL-receptor class B 33.					
FT	4213	4249	EGF-like 13.					
FT	4249	4290	EGF-like 14.					
FT	4286	4326	EGF-like 15.					
FT	4321	4363	EGF-like 16.					
FT	4392	4430	EGF-like 17.					
FT	32	45	By similarity.					
FT	39	58	By similarity.					
FT	52	69	By similarity.					
FT	77	90	By similarity.					
FT	84	103	By similarity.					
FT	97	113	By similarity.					
FT	159	169	By similarity.					
FT	165	178	By similarity.					

QY 93 -----QT--R-Q-Q-----AY--R-----DK- 102
Db 900 PKRWLCGANDCGSNEDESNOTCTARTCQADQPSGNGRCIPATWLCDDREDDCGDQDEV 959
QY 103 -----LAQ-----Q----- 106
Db 960 ASCEPFTCEPLTOFICKSGRCISKNWHCDTDDCGDRSDRDEVCVHSCLDQDFCSCGRCI 1019
QY 107 --H-A-----AAAA--A----- 114
Db 1020 PGHWACDGDNDGDFSETHINCKEARSAGCIGNEQCRPDGNCIPDLWRCDGEKDC 1079
QY 115 -----A-A-AT-----N-----Q----- 120
Db 1080 EDGSEKGCNGTIRLCDHKTFCSCRTGRCINNAWVCGDGDVDCEDQSDDEEDCDSFLCGPE 1139
QY 121 -----Q-----GSAK-----NG----- 127
Db 1140 KYPCANDTSVCLQPEKLCNGRKDCPDGSDGDLCDCECLNNGCSNHCSVVPGRGIVCSC 1199
QY 128 -E-----N-T-----A-----A-NGE-----E 135
Db 1200 PEGHQLKDNRTCEIVDYCASHLRCQVCEQQHMKVCSCYEGWALGTGDESCTSVDSFE 1259
QY 136 -----TD-----M-----N-----E--- 150
Db 1260 APTIISRHEIRIDLHKGDSLLVPLGRNTIALDPHFQNSLLYTDVVEDRIYRKLSE 1319
QY 137 --G-A--H-----T--IA-N-----N-----H 145
Db 1320 SGVSAIEVVVEHGLATPEGLTVDWIAIGNIYWDNSLDQTEVSKDGLSLRATLIAGAMEH 1379
QY 146 -----TD-----M-----M-----E--- 150
Db 1380 PRAIALDPYGLFTWDANPPRIESASMSGAGRTIYKDMKTGAWPNGLTVDHFERI 1439
QY 151 V-----DG-D-VEI-----P-----S-----N-KA----- 162
Db 1440 VMTDARSDAIYSAFYDGTNMBIEIRGHEVLSHPFAVSLYGEVYVTDWRTNLTAKANKWT 1499
QY 163 --VV-----L-----RG-----H- 168
Db 1500 GONVSVIOKTSQAPFDLOIYHPSROPAPNCAANEGRGFCSHLCLINHNRSACACPHL 1559
QY 169 -----E-----SE-----V----- 172
Db 1560 MKLSSDKTCYEMKKFLYARSEIRGVIDIDNPYNFIPTFTVDDDDVAVDFDASEER 1619
QY 173 -----FI----- 174
Db 1620 LYWTDIKTQITITRAFNGTGLETVISRDIGSLAVDMSVRLNLYWISSEFDEFTQINVAR 1679
QY 175 -----C-A-----W-----N--PV--S 181
Db 1680 LDGSLKTSIIGHIDKPCQCLAAHPVRGKLYWTDGNTINMANMGNSKILFQNOKEPVGLS 1739
QY 182 -D-----LQ-V-SGSG-----D-----S-----TAR-I-----W--NL-- 198
Db 1740 IDYVENKLYWISSGNGTINRCLDGNLELVIESMKEELTKATALTITMDKMLWADQNLAQ 1799
QY 199 -----S-----E-----NS-----TS----- 204
Db 1800 LGTCNKRDGRNPISLANKTSVHVHMKVYDKEAQGNSCOVNVNGCSQLCLPSETTRTC 1859
QY 205 -----G-----P----- 206
Db 1860 MCTVGYLQKNRMSCOGIESFLMYSVHEGIRGIPLEPRKVDALMPISGAFAVGDIFHA 1919
QY 207 -----T-----Q-----LV--L-R-----H-----C- 214
Db 1920 ENDTIYTDWGLNKSIRAKDQTKVEDVTNGLRVEGIAVDWIAIGNIYWTDHGNLIEV 1979
QY 215 -----I-----R-----E-G-----GQ-----D-----V-- 222

Db 1980 ARLNGSFRYVILISQGLDQPRSIATVPEKGFLEWTEMGQVPCIGKARLDGSEKVMIVSVGI 2039
QY 223 --P--S-----NK-----D--V--T-----S-LD-----W- 233
Db 2040 TWPNGISIDYEENKLYWCARDARKIERIDLTGANREVLSSGNSVDLFSVAVFAGIYWS 2099
QY 234 -----N-S-----E-GT----- 238
Db 2100 DRAHANGSVRRGHKNDAETVTWRTGLVNLKEIKIFNRVREKGTNVCAKENGCCQQLCL 2159
QY 239 -----LLA--T-----G-----S----- 244
Db 2160 YRGNSRRTCACAHGYLAGDVTCLRHGEGYLLSGRTILKSIHLSDETNLNLSPPRYENPN 2219
QY 245 Y-----D-----G--F--A-----R--I----- 251
Db 2220 YFNIIALAFYNORREGTNRIFYSDAHFNGIQLIKDNWEDROVIVENGVSGEGLAYHRA 2279
QY 252 -----W-----TK--D--G-----NL----- 258
Db 2280 WDTLYWTSSTSSITPHTVDQTRPGAI DREAVITMSEDDPHVLADECONLFWTNWNE 2339
QY 259 -----A--ST-----LG-----Q-H- 265
Db 2340 QHPSIMRATLTGKNAHVVVSTDLTPNGLTIDHRAEKLYFSDGSLGKIERCEYDGSQRHV 2399
QY 266 --K-GP-----IF-ALKW-----NK-----K----- 277
Db 2400 IVKSGEGTFLSAVDSYIFWS--DMGRRAILRSNKYTGGETKILRSDI PHQPGIITAVAN 2458
QY 278 -----G-----N-----F-- 280
Db 2459 DTNSCELSPCALLGGCHDLCLLTDPGRVNCSCRGDRVLLANNRCVTKNSSCNIYSEFEC 2518
QY 281 -----I-----L-----S-AG-----V----- 286
Db 2519 GNGDCVYVLTCDGIPHCKDKSDEKLLYCCENSRSGFKPCYNRRRCVPHGKLCGTNDGC 2578
QY 287 -----D-K--T--T-I----- 291
Db 2579 DSSDELDCVSTCSTVEFRCADGTICIPRSARCNQNMDCSDASDEKGCNNTDCTHFYKLG 2638
QY 292 -----I-----W-----D-----AH-----T 297
Db 2639 KSTGFIRCNSTSLCVLPSWICDGSNDGDYSDCLKPCVQNKHKEBENYFCGSPGRCILNT 2698
QY 298 -----G-E-----A-----K----- 301
Db 2699 WVCDOGKOCEDGLDELHDCSSCWNOFACSVKKICKSHWICDGEDDCGSDLSDESITCGA 2758
QY 302 -----Q-----Q-----F----- 304
Db 2759 VTCAADMFCQSHACVPOHWCIDGERDCPDGSDLSAGCAPNNTCDENAFMCHNKVCI 2818
QY 305 P-F--H-----S-----AP-----ALD-----VD--WQ-----S-- 318
Db 2819 PKQFVCDHDDDCGSDGSEFLQCYRCQCGPEERCA--DGRCLVNTLWQCGDPCDPSSDE 2877
QY 319 --N-----NT--FASC-----STD-MC-I-----HV--C--KL 336
Db 2878 APINPRCSAEHSNCSFFM--CKNGRCIPS--DGLCDIRDGSDGSDTNGHINECLSKKI 2935
QY 337 -G-QD-R--PI-----KT-----F--Q-----G----- 347
Db 2936 SGCSQDQDLPVSYKCKWPGFQLKDDGKTCVDIDECSSGFPSCQOCINTYGYKCHCAE 2995
QY 348 -----H-----H-----T-----N-----E 351
Db 2996 GYETQDPNPGCRSLSDPEPFLILADQHEIRKISTDGSNTLLKQGLNNVIALDFDYREE 3055
QY 352 -----I-----V-NA-----I-K--W-D----- 358

Db 3056 FIYWIDSSRPNGSRINRMCLNGSDIKVHVHTAPVNALVDWIGKNLYWSDTEKRIIEVSK 3115
QY 359 -----PT-----GNLL-ASC-----S- 368
Db 3116 LNLGYPTVLVSKRLKFPRLSLDPAGNLYWIDCCEYPHIGRVGMDGTNQSVIETKISR 3175
QY 369 -----D-----M-----TL-K-I-W----- 376
Db 3176 PMALTIYVNHLYWADENHIEFSNMDSHRHKVPNQDIPGVIALTLFEDIYWDGKTK 3235
QY 377 ---S-MK-----Q-D-----N-----C-----V 384
Db 3236 SILRVHKTSGADRLSLNLSWHAITDIQVYHSVRQPDVSKHLCTVWNGCSHLCLLGPCKT 3295
QY 385 H-----D-----L----- 387
Db 3296 HTCACPTNYLAADNRCTLSNCTASQFRCKTDKIPFWWKDVTVDGCGDSDEPDDCPEF 3355
QY 388 -----Q-----A-----H-----NKE-I----- 394
Db 3356 KCQPRFOCGTGLCALPAFICDGENDCGNSDELNCDTHVCLAGQFKCKNKKCIPVNL 3415
QY 395 -----Y-----T-----I-K-W-----SP-----T 402
Db 3416 CNGQDDCDEDEKDCPENSCSPDYFOCKTKTKGICSLWVCEDEPDCCADASDEANCDKKT 3475
QY 403 -GP-----G-----T-N-----N-P----- 409
Db 3476 CGPHEFOCKNNNCIPDHWRCNDQNSDNDCKFQCTLKDPLCSNGDCVSSRFWCD 3535
QY 410 -----N-----A-----N----- 412
Db 3536 GEFDCADGSDKNCETSKDOFQCSNGQCLSAKWCKGHEDCKYGEDEKNCBPAPFVCS 3595
QY 413 ---LMLA-----SAS-----F----- 420
Db 3596 SSEYMCASGGCLSLAKNGEPCDVGSDMDCKVECKEDQFOCKNKAYCIPIRWLCDGI 3655
QY 421 ---D-S---T-----V-R-----L-W-----D-V----- 429
Db 3656 YDCVGSDEEFCGRGSGICRDDEFLCNLSLKLHFWVCDGEDDCGDSDEAPDMCVKFLC 3715
QY 430 -----DR-----GI-----CI 435
Db 3716 PPTRPYRCNRDRI CLQLEKICNGINDCGDSDDEHSCGKLSLSPCKCKDEFTCSNRNCI 3775
QY 436 -----HT----- 437
Db 3776 PMLQCDLDDCGGSDGDEOQGLKTPIEHTCENNGNPGDDAYCNQIKTSVFCRCXPFQR 3835
QY 438 -----L-T-----K----- 440
Db 3836 NMKGRECADLNECLLFGICSHHCLNTRGSKVCQDNQFQEKNNCTAKGSEDOALYIAND 3895
QY 441 -----HQ-----E-----P-V-Y----- 446
Db 3896 TDILGFVYPFNYSGHQOISHVHENSRITGMVHYQRNVIVWSTQFNPGGIFYKIMIDARE 3955
QY 447 ---S-----VA-----RS----- 451
Db 3956 KRQANSGLICPEKPRPDIAVDWAGNVYVTHSRMHWFYSYYTHWTSLSYSINVGQLNG 4015
QY 452 -----P-----DG----- 454
Db 4016 PNCTRLTNMAGEPYAIANPFRGMMYWTVIGDHSHEEAAAMDGTLRRVLVQKNLQRP TG 4075
QY 455 -----R-Y-----LA---S---GS----- 461
Db 4076 LTVDFGERIYWADFELSIIGSVLYDGSPPVSVSSKQGLLPHPRIDVFEDIYAGPKN 4135
QY 462 -----F-----DK-----CV 466
Db 4136 GIFRVQKFGHGSVEVLALGVDTKTSILVSHRYKQLNLPNCLDLSCDFLCLLNPSGATCI 4195

QY 467 -----H-----I-----W-----N-T- 471
Db 4196 CPEGKYMMNGTCHDDSLDDSKLTCENGRGCIILNEKGLDLRCHCWPSYSGRCVNHCSN 4255
QY 472 --Q-----T-G-----AL-----V-----H-----S-----Y 480
Db 4256 YCQNGGTCTPSTIGRPTCICALGFTGPNCGKAVCEDSCHNGSCVVTAGNQPYCHCOADY 4315
QY 481 ---R-----G-----T-----GG----- 485
Db 4316 TGDRCQYVCHHYCVNSECTIGNDGSVEVCPTRYEGPKCEIDKCVRCHGHCIIKNON 4375
QY 486 --IF-----E-----VC---W----- 491
Db 4376 EDIFCNCNTNGKIASQCQCDGYCYNGGTQCLDPETSI P V C V C S T N N S G T Q C E R P A P K S S K 4435
QY 492 -----C-----VLDLRK 514
Db 4436 SEHISTRSTAIIVPLVLLVTLVTLVIGLVVCKRRTKTIRROPIINGINVEIGNPSY 4495
QY 492 NA-----AG-----D-K---VG---A-----SA-----SD-----G--- 505
Db 4496 NMVEVDHDSGDLLEPSFMIDPVKSRYIGGSSAFKLPHTAPPIVYLSNLSLKGPLTFGPT 4555
QY 506 --S--V-----C-----VLDLRK 514
Db 4556 NYSNPFYAKLYMDGQNCNRNSLASV-DERK 4583

Search completed: January 3, 2005, 15:26:03
Job time : 135.333 secs